

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2021년 2월 25일 (25.02.2021) **WIPO | PCT**



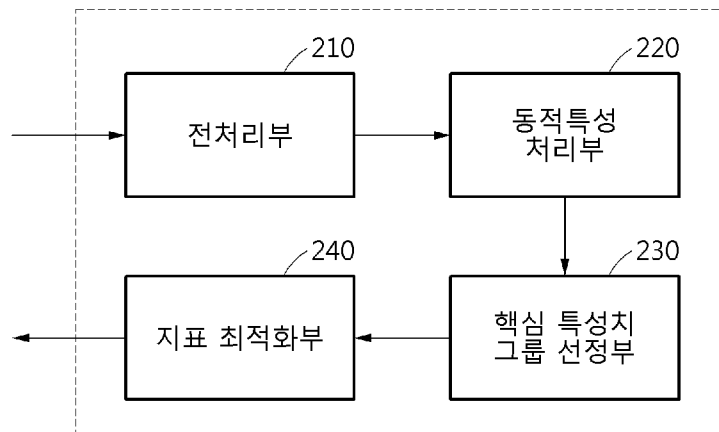
(10) 국제공개번호
WO 2021/034106 A1

- (51) 국제특허분류: **G06N 3/08** (2006.01) **G01W 1/10** (2006.01)
G06N 3/04 (2006.01)
- (21) 국제출원번호: PCT/KR2020/011052
- (22) 국제출원일: 2020년 8월 19일 (19.08.2020)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2019-0101069 2019년 8월 19일 (19.08.2019) KR
- (71) 출원인: 고려대학교 산학협력단 (**KOREA UNIVERSITY RESEARCH AND BUSINESS FOUNDATION**)
[KR/KR]: 02841 서울시 성북구 안암로 145, Seoul (KR).
- (72) 발명자: 오윤영 (**OH, Yun Yeong**): 47711 부산시 동래구 중앙대로 1473번길 14-12, 1527호, Busan (KR). 윤성택 (**YUN, Seong Tack**): 02794 서울시 성북구 월곡로 14길 26, 108동 1003호, Seoul (KR).
- (74) 대리인: 김연권 (**KIM, Youn Gwon**): 05836 서울시 송파구 법원로 127 문정대명벨리온 15층 1501호 및 1502호, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: DEVICE FOR OPTIMIZING TRAINING INDICATOR OF ENVIRONMENT PREDICTION MODEL, AND METHOD FOR OPERATING SAME

(54) 발명의 명칭: 환경 예측 모델의 훈련지표 최적화 장치 및 그 동작방법

200



210 ... Preprocessing unit

220 ... Dynamic feature processing unit

230 ... Key feature group selection unit

240 ... Indicator optimization unit

(57) Abstract: The present invention relates to a device for optimizing a training indicator of an environment prediction model, and a method for operating same. A device for optimizing a training indicator according to one embodiment comprises: a preprocessing unit which composes a base data set of environment measurement data; a dynamic feature processing unit which identifies and extracts dynamic features of the composed base data set through multi-resolution wavelet analysis and a dimension reduction technique; a key feature group selection unit which identifies and evaluates a driving force for the environment measurement data on the basis of the extracted dynamic features, and selects a key feature group in response to the evaluation result; and an indicator optimization unit which receives the selected key feature group and the environment measurement data as inputs, and controls a plurality of training indicator values corresponding to an environment prediction model.



WO 2021/034106 A1

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(57) 요약서: 본 발명은 환경 예측 모델의 훈련지표 최적화 장치 및 그 동작방법에 관한 것으로서, 일실시예에 따른 훈련지표 최적화 장치는 환경 계측 자료에 대한 기반데이터셋을 구성하는 전처리부와, 다해상도 웨이블릿 해석과 차원 축소 기법을 통해 구성된 기반데이터셋에 대한 동적 특성을 식별 및 추출하는 동적 특성 처리부와, 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별 및 평가하고, 평가 결과에 대응하여 핵심 특성치 그룹(key features group)을 선정하는 핵심 특성치 그룹 선정부 및 선정된 핵심 특성치 그룹과 환경 계측 자료를 입력으로 수신하여 환경 예측 모델에 대응되는 복수의 훈련지표 값을 제어하는 지표 최적화부를 포함한다.

명세서

발명의 명칭: 환경 예측 모델의 훈련지표 최적화 장치 및 그 동작방법

기술분야

- [1] 본 발명은 환경 예측 모델의 훈련지표 최적화 장치 및 그 동작방법에 관한 것으로서, 보다 상세하게는 통합환경 모니터링을 위한 딥러닝 알고리즘의 훈련지표를 최적화하는 기술적 사상에 관한 것이다.

배경기술

- [2] 고전적인 RNN(recurrent neural network; 순환신경망)의 기울기 소실(vanishing gradient)과 장기기억의존성 문제를 해결하기 위해 고안된 LSTM(long-short term neural network; 장단기 기억네트워크)은 셀 스테이트(cell state)와 망각 게이트(forget gate)를 이용한 유연한 가중치 조절을 통해 장기기억 의존성(long-term memory dependencies) 문제를 매우 효율적으로 다룰 수 있다.
- [3] LSTM은 게이트(gate)라고 불리는 시그모이드층(sigmoid layer)과, 점 단위 조합을 갖는 일종의 가중치 조절 함수군(functional group)을 이용하여 셀 스테이트를 개선 및 제어하며, 특히 망각 게이트를 이용한 선별적 업데이트를 통해 기존의 RNN에 비해 매우 유연한 적응성을 갖는 것을 특징으로 하고 있다.
- [4] LSTM에는 다양한 변종(variation)들이 존재하며, 대표적인 예시로써 Peenhole connection model (Gers and Schmidhuber, 2000), Gated recurrent unit model (Cho et al., 2014), Depth gated RNN (Yao et al., 2015) 등을 들 수 있다.
- [5] 그러나 기존의 많은 연구들과 개선된 모델들에도 불구하고, 데이터의 특성에 따라 어떤 LSTM 모델을 사용하는 것이 효율적이며, 또한 동일한 데이터에 대해서 같은 LSTM을 적용시에도 데이터의 어떤 특성(feature)을, 어떠한 훈련 조건(training conditions or options) 하에서 훈련시키는 것이 최적의 예측능 발휘에 효과적 인지를 나타내는 '훈련조건 최적화(optimization conditions for model training) 문제'에서는 공인된 정량화 기법이 없으며, 개별 사용자들의 경험에 따른 시행 착오 기반의 방법(trial-error based methods)에 의존하고 있다.

발명의 상세한 설명

기술적 과제

- [6] 본 발명은 데이터의 특성과 모델링 목적에 따른 최적 훈련조건을 사전에 결정할 수 있는 훈련지표 최적화 장치 및 그 방법을 제공하고자 한다.
- [7] 또한, 본 발명은 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 훈련시간 단축, 핵심 특성에 대한 집중학습 및 과적합을 방지할 수 있는 훈련지표 최적화 장치 및 그 방법을 제공하고자 한다.
- [8] 또한, 본 발명은 최적 훈련조건을 사전에 결정함으로써, 기존의 LSTM 모델들에 대해서도 입력자료에 따른 실시간 모델 업데이트를 통해 예측능을

보다 더 향상시킴과 동시에 최적훈련지표 선정을 위한 정량적 근거를 제시하여 다양한 환경자료들을 이용한 통합환경모니터링, 해석, 예측 및 대응시스템의 구축에서 보다 효율적인 환경 예측 모델을 제공할 수 있는 훈련지표 최적화 장치 및 그 방법을 제공하고자 한다.

과제 해결 수단

- [9] 일 실시예에 따른 훈련지표 최적화 장치는 환경 계측 자료에 대한 기반데이터세트를 구성하는 전처리부와, 다해상도 웨이블렛 해석과 차원 축소 기법을 통해 구성된 기반데이터세트에 대한 동적 특성을 식별 및 추출하는 동적 특성 처리부와, 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별 및 평가하고, 평가 결과에 대응하여 핵심 특성치 그룹(key features group)을 선정하는 핵심 특성치 그룹 선정부 및 선정된 핵심 특성치 그룹과 환경 계측 자료를 입력으로 수신하여 환경 예측 모델에 대응되는 복수의 훈련지표 값을 제어하는 지표 최적화부를 포함할 수 있다.
- [10] 일측에 따르면, 환경 계측 자료는, 실시간으로 계측되는 수문-환경 시계열 자료를 포함하고, 수문-환경 시계열 자료는, 수문 기상 데이터, 하천 수위 데이터, 지하 수위 데이터, 수질 데이터, 온도 데이터, EC 데이터, 동위원소비율 데이터, 토양 가스 데이터 및 미세먼지 데이터 중 적어도 하나 이상의 환경 데이터를 포함할 수 있다.
- [11] 일측에 따르면, 전처리부는, 환경 계측 자료의 데이터세트의 관측항목과 관측시간 해상도에 따른 데이터 매트릭스를 기반데이터세트로 구성 및 정렬하고, 정렬된 기반데이터세트에 대한 시간영역 해상도 또는 시간관측 간격별 결측자료를 보간 처리 하고, 보간 처리된 기반데이터세트의 자료를 노이즈 필터링하며, 노이즈 필터링된 결과를 표준화 및 정규화할 수 있다.
- [12] 일측에 따르면, 동적 특성 처리부는, 구성된 기반데이터세트에 대한 시간영역 해상도에 따른 다해상도 웨이블렛 해석을 통해 시간-주파수 영역에 대한 웨이블렛 에너지 분포 데이터를 도출하고, 도출된 웨이블렛 에너지 분포 데이터에 대하여 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군을 선정할 수 있다.
- [13] 일측에 따르면, 동적 특성 처리부는, 선정된 잠재 환경 구동력 후보군의 시간영역 해상도별 시간변화에 따른 변동특성을 추출하고, 추출된 변동특성에 기초하여 동적 특성을 추출 및 정량화할 수 있다.
- [14] 일측에 따르면, 차원 축소 기법은, 주성분/독립성분 분석(principle/independent component analysis; PCA/ICA), 시계열 요인분석(time series factor analysis, TSFA), 경험적모드분해(empirical mode decomposition; EMD) 및 다해상도상태공간모델(multiresolution state-space model; MRSSM) 중 적어도 하나 이상의 기법을 포함할 수 있다.
- [15] 일측에 따르면, 핵심 특성치 그룹 선정부는, 잠재 환경 구동력 후보군의 잠재

환경 구동력과, 환경 계측 자료 간 다해상도 상관성을 판단하되, 잠재 환경 구동력과 관측자료간 시간지연 및 위상변화를 반영한 상관성 판단을 수행하고, 수행된 상관성 판단 결과에 기초하여, 잠재 환경 구동력과 관측자료간 최대상관스케일을 선정할 수 있다.

- [16] 일측에 따르면, 핵심 특성치 그룹 선정부는, 잠재 환경 구동력과 관측자료간 웨이블릿 에너지 비율과 선정된 최대상관스케일의 상관관계를 이용한 구동력을 식별하고, 선정된 최대상관스케일의 결합에너지비율과 차원축소모델의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 평가된 상대적 기여도에 기초하여 핵심 특성치 그룹을 선정할 수 있다.
- [17] 일측에 따르면, 핵심 특성치 그룹 선정부는, 잠재 환경 구동력 후보군 및 핵심 특성치 그룹 중 적어도 하나 이상을 이용하여 훈련되는 사전 조정된 LSTM 네트워크(well-tuned LSTM networks)를 구축하고, 사전 조정된 LSTM 네트워크를 이용하여 잠재 환경 구동력을 검증할 수 있다.
- [18] 일측에 따르면, 지표 최적화부는, 핵심 특성치 그룹과 환경 계측 자료를 입력으로 하는 장단기기억네트워크(Long-short Term Memory Network) 모델을 구축하고, 핵심 특성치 그룹의 시간-주파수 영역에 기초하여 복수의 훈련지표 값을 사전 정량화(pre-quantification)할 수 있다.
- [19] 일측에 따르면, 지표 최적화부는, 환경 계측 자료로부터 측정되는 관측값과, 장단기기억네트워크 모델로부터 예측되는 예측값의 복합모델검증지표에 기초한 잔차검증 및 잔차의 다해상도해석에 기반하여 적어도 하나 이상의 예측 모델을 선정하고, 선정된 적어도 하나 이상의 예측 모델 중 어느 하나의 예측 모델 또는 둘 이상의 예측 모델이 결합한 결합 예측 모델(combined prediction model)에 기초하여 사전 정량화된 복수의 훈련지표를 정량화할 수 있다.
- [20] 일측에 따르면, 지표 최적화부는, 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 이용하여 원 자료의 특성에 따른 훈련지표 최적화 모형을 구성할 수 있다. 예를 들면, 원 자료는 환경 계측 자료일 수 있다.
- [21] 일측에 따르면, 복수의 훈련지표 값은, 훈련기간(training period, T), 최소 배치 사이즈(minibatch size, mbs), 은닉층(hidden layers, HL)의 개수 및 최적 반복회수(epochs, E) 중 적어도 하나의 값을 포함할 수 있다.
- [22] 일 실시예에 따른 훈련지표 최적화 방법은 전처리부에서, 환경 계측 자료에 대한 기반데이터셋을 구성하는 단계와, 동적 특성 처리부에서, 다해상도 웨이블릿 해석과 차원 축소 기법을 통해 구성된 기반데이터셋에 대한 동적 특성을 식별 및 추출하는 단계와, 핵심 특성치 그룹 선정부에서, 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별 및 평가하고, 평가 결과에 대응하여 핵심 특성치 그룹(key features group)을 선정하는 단계 및 지표 최적화부에서, 선정된 핵심 특성치 그룹과 환경 계측 자료를 입력으로 수신하여 환경 예측 모델의 복수의 훈련지표 값을 제어하는 단계를 포함할 수 있다.
- [23] 일측에 따르면, 동적 특성을 식별 및 추출하는 단계는, 구성된

기반데이터세트에 대한 시간영역 해상도에 따른 다해상도 웨이블릿 해석을 통해 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 도출하고, 도출된 웨이블릿 에너지 분포 데이터에 대하여 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군을 선정하는 단계 및 선정된 잠재 환경 구동력 후보군의 시간영역 해상도별 시간변화에 따른 변동특성을 추출하고, 추출된 변동특성에 기초하여 동적 특성을 추출 및 정량화하는 단계를 더 포함할 수 있다.

- [24] 일측에 따르면, 핵심 특성치 그룹을 선정하는 단계는, 잠재 환경 구동력 후보군의 잠재 환경 구동력과, 환경 계측 자료 간 다해상도 상관성을 판단하되, 잠재 환경 구동력과 관측자료간 시간지연 및 위상변화를 반영한 상관성 판단을 수행하고, 수행된 상관성 판단 결과에 기초하여, 잠재 환경 구동력과 관측자료간 최대상관스케일을 선정하는 단계 및 잠재 환경 구동력과 관측자료간 웨이블릿 에너지 비율과 선정된 최대상관스케일의 상관관계를 이용하여 구동력을 식별하고, 선정된 최대상관스케일의 결합에너지비율과 차원축소모델의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 평가된 상대적 기여도에 기초하여 핵심 특성치 그룹을 선정하는 단계를 더 포함할 수 있다.

- [25] 일측에 따르면, 복수의 훈련지표 값을 제어하는 단계는, 핵심 특성치 그룹과 환경 계측 자료를 입력으로 하는 장단기기억네트워크(Long-short Term Memory Network) 모델을 구축하고, 핵심 특성치 그룹의 시간-주파수 영역에 기초하여 복수의 훈련지표 값을 사전 정량화(pre-quantification)하는 단계와, 환경 계측 자료로부터 측정되는 관측값과, 장단기기억네트워크 모델로부터 예측되는 예측값의 복합모델검증지표에 기초한 잔차검증 및 잔차의 다해상도해석에 기반하여 적어도 하나 이상의 예측 모델을 선정하고, 선정된 적어도 하나 이상의 예측 모델 중 어느 하나의 예측 모델 또는 둘 이상의 예측 모델이 결합한 결합 예측 모델(combined prediction model)에 기초하여 사전 정량화된 복수의 훈련지표를 정량화하는 단계 및 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 이용하여 최적 훈련지표모형을 구성하는 단계를 더 포함할 수 있다.

- [26] 일측에 따르면, 복수의 훈련지표 값은 훈련기간(training period, T), 최소 배치 사이즈(minibatch size, mbs), 은닉층(hidden layers, HL)의 개수 및 최적 반복회수(epochs, E) 중 적어도 하나의 값을 포함할 수 있다.

발명의 효과

- [27] 일실시예에 따르면, 데이터의 특성과 모델링 목적에 따른 최적 훈련조건을 사전에 결정할 수 있다.
- [28] 일실시예에 따르면, 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 훈련시간 단축, 핵심 특성에 대한 집중학습 및 과적합을 방지할 수 있다.
- [29] 일실시예에 따르면, 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 입력자료에 대한 실시간 모델 업데이트를 통해 예측능을

보다 더 향상시킴과 동시에 최적훈련지표 선정을 위한 정량적 근거를 제시함으로써 다양한 환경자료들을 이용한 통합환경모니터링, 해석, 예측 및 대응시스템 구축에 보다 효율적인 환경 예측 모델을 제공할 수 있다.

도면의 간단한 설명

- [30] 도 1은 일실시예에 따른 환경 예측 모델의 훈련지표를 최적화하는 훈련지표 최적화 장치의 적용 개념을 설명하기 위한 도면이다.
- [31] 도 2는 일실시예에 따른 훈련지표 최적화 장치의 구성을 설명하기 위한 도면이다.
- [32] 도 3a 내지 도 3h는 일실시예에 따른 훈련지표 최적화 장치에서 토양 가스를 환경 계측 자료로 수신하여 동작하는 예시를 설명하기 위한 도면이다.
- [33] 도 4a 내지 도 4f는 일실시예에 따른 훈련지표 최적화 장치에서 하천/ 지하수 수위(RWL/GWL) 및 전기 전도도(RWEC/ GWEC) 계측 자료를 환경 계측 자료로 수신하여 동작하는 예시를 설명하기 위한 도면이다.
- [34] 도 5a 내지 도 5c는 일실시예에 따른 최적훈련지표모형을 구성하는 예시를 설명하기 위한 도면이다.
- [35] 도 6은 일실시예에 따른 훈련지표 최적화 방법을 설명하기 위한 도면이다.

발명의 실시를 위한 형태

- [36] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.
- [37] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [38] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만, 예를 들어 본 발명의 개념에 따른 권리 범위로부터 이탈되지 않은 채, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [39] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는

것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 표현들, 예를 들어 "~사이에"와 "바로~사이에" 또는 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

- [40] 본 명세서에서 사용한 용어는 단지 특정한 실시예들을 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [41] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [42] 이하, 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [43] 도 1은 일실시예에 따른 환경 예측 모델의 훈련지표를 최적화하는 훈련지표 최적화 장치의 적용 개념을 설명하기 위한 도면이다.
- [44] 도 1을 참조하면, 참조부호 100은 일실시예에 따른 환경 예측 모델의 훈련지표 최적화 장치를 설명하기 위한 것으로, 환경 예측 모델은 MR-LSTM(multi-resolution long-short term neural network) 모델일 수 있다.
- [45] 참조부호 100에 따르면, 일실시예에 따른 훈련지표 최적화 장치는 실시간으로 계속되는 환경 예측 자료들 및 환경변수 예측자료들에 대해, 다해상도 시간-주파수 영역의 해석방법을 이용하여 관측값 변화의 원인이 되는 핵심적인 환경구동력들(key environmental drivers)과 그 효과들을 보다 체계적으로 분리 및 식별하고, 기여도에 대한 정량적 평가(evaluation)를 수행할 수 있다.
- [46] 또한, 훈련지표 최적화 장치는 정량적 평가를 기반으로 하여 실제 관측값과 관측값에 내재되어 있는 환경요인을 동시에 이용하는 LSTM 심층신경망(deep learning neural network)을 구성할 수 있다.
- [47] 또한, 훈련지표 최적화 장치는 최적 학습을 위한 특성치 선정 및 예측의 인과성에 대한 정량적인 해석의 근거를 제시하고, 평가 및 예측의 대상이 되는 환경예측 자료 및 핵심환경요인에 대한 시-공간적 변동 특성에 대해 보다 집중적인 학습을 가능하게 하여, 최적훈련 조건을 사전에 선정하거나 사후 개선 시 훈련시간 단축, 모델의 적합도 및 예측능 향상을 기대할 수 있다.
- [48] 또한, 훈련지표 최적화 장치는 특정 환경예측 자료들의 종류에 구속받지 않고

실시간으로 계측되는 다양한 목적변수들의 분리, 식별, 평가 및 예측에 범용적으로 활용 가능하며, 특히 통합환경모니터링 해석 및 관리시스템에서의 위험 신호의 조기 식별 및 탐지, 특정 목적 환경변수와 복합 환경요인에 대한 취약성/위해성 평가(vulnerability/risk assessment) 및 이를 기반으로 하는 실시간 대응 시스템(real time response system)의 구현과 그 운용 있어 효율성 향상에 크게 기여할 수 있다.

- [49] 예를 들면, 환경 계측 자료들 및 환경변수 계측자료들은 수문-환경시계열 자료(수문기상자료, 하천 및 지하수 수위, 수질, 온도, EC자료(토양/ 하천수/ 지하수 전기전도도 자료), 동위원소비율), 토양가스자료(CO₂, NO₂, CO, NO, SO₂, Rn) 및 미세먼지 자료(PM₁₀, PM_{2.5})를 포함할 수 있다.
- [50] 구체적으로, 훈련지표 최적화 장치는 환경 계측 자료들 및 환경변수 계측자료들의 구조적 특성과 인과성의 근거를 차원 축소모형 (예를 들면 다해상도상태공간모형)을 통해 확보할 수 있다.
- [51] 참고로, 일반적인 요인 분석(factor analysis)은 다변량데이터의 차원축소를 통한 주요요인의 정량화 기법에 해당한다. 또한, 상태공간모형 중 동적요인분석(dynamic factor analysis)은 시계열분석과 요인분석의 결합을 통한 관측자료에 내재하는 공통요인을 추출하고 분석할 수 있다. 또한, 시계열 분석(time series analysis)은 시간 순서로 연속적으로 관측되는 자료들에 대한 추계적인(stochastic) 분석 기법의 총칭으로, 광범위한 분야에 적용되는 보통명사로 해석될 수 있다.
- [52] 한편, 일실시에에 따른 훈련지표 최적화 장치는 웨이블릿 해석기반의 다해상도상관성 분석을 통해 잠재구동력과 관측자료의 시간-주파수영역의 유사성을 정량화함으로써 환경 구동력을 구체적으로 식별하고, 영향력을 정량적으로 평가할 수 있다.
- [53] 예를 들면, 웨이블릿 해석은 하기 수학적 1 내지 4를 통해 도출되는 웨이블릿 필터링된 관찰 함수에 의해 관측값을 산출할 수 있다.

[54] [수학적식1]

[55]

$$\psi_{a,b}(t) = a^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right)$$

[56] 여기서, ψ 는 웨이블릿 모 함수(wavelet mother function), a는 스케일

계수(scale coefficient), b는 전이 계수(translation coefficient)를 나타내고, 스케일 계수는 대표 주기(represent period)의 사이즈를 갖는 계수이며, 변형 계수는 시간 축(time axis)에서의 이동 위치에 대응되는 계수일 수 있다.

[57] [수학적식2]

[58]

$$f(t) = \overbrace{\sum_{j,k \in \mathbb{Z}} c_{j,k} \Psi_{j,k}(t)}^{\text{CWT}} \approx \underbrace{\sum_{j \in \mathbb{Z}} c_{0,j} \Psi_{0,j}(t)}_{\text{Approximations}} + \underbrace{\sum_{k \geq 0} \sum_{j \in \mathbb{Z}} d_{k,j} \Phi_{k,j}(t)}_{\text{Details}} \quad \text{DWT (dyadic)}$$

[59] 여기서, Ψ 는 긴 스케일(long scale) 및 저주파수(low frequency)에 대응되는 스케일링 함수(scaling function), Φ 는 짧은 스케일(short scale) 및 고주파수(high frequency)에 대응되는 웨이블릿 함수(wavelet function)일 수 있다.

[60] [수학식3]

[61]

$$f_i^w(t) = \sum_{j=1}^k \alpha_{i,j} F_j(t) + \mu_i(t) + \varepsilon_i(t)$$

[62] 여기서, $f_i^w(t)$ 는 i번째 웨이블릿 필터링된 관측변수(wavelet filtered observation)($1 \leq i \leq t$), $F_j(t)$ 는 j번째 잠재 공통인자(latent common factor)($1 \leq j \leq k$), $\alpha_{i,j}$ 는 동적요인부하량(dynamic factor loading), $\mu_i(t)$ 는 i번째 상수 레벨 파라미터(constant level parameter), $\varepsilon_i(t)$ 는 잔차항(residual term)에 대응되는 특정인자(specific factor)일 수 있다.

[63] [수학식4]

[64]

$$F_j(t) = \sum_{p=1}^n F_j(t-p) + \eta_j(t)$$

[65] 여기서, $F_j(t)$ 는 j번째 잠재 공통인자(latent common factor), p는 최적 자기회귀 차수(optimal autoregressive order), $\eta_j(t)$ 는 백색잡음(white noise)일 수 있다.

[66] 본 명세서에서 사용하는 웨이블릿 해석(wavelet analysis)이라 함은, 시계열자료의 웨이블릿 변환(wavelet transform)을 이용한 시간-주파수영역의 다해상도 분석 기법으로 해석될 수 있다. 웨이블릿 해석은 관측자료를 다양한 시간-주파수 영역으로 분해하는데 특화되어 있다. 즉, 웨이블릿 해석은 시간 영역과 주파수 영역을 동시에 분석하는 방법의 하나로, 연속 신호와 이산 신호에 모두 적용할 수 있으며, 결함 진단에 널리 적용될 수 있는 기법이다.

- [67] FFT(fast fourier transform)는 측정 데이터 내에서 정보가 시간 평균이 되므로, 시간 구간에서의 정보를 잃어버리는 단점이 있다. 따라서 웨이블릿 변환은 특히 시간에 대한 결함 주파수가 변화하는 비정상 신호(non-stationary signal)나 과도 신호(transient signal)의 분석에 유용하다. 일반적인 단순 푸리에 변환(STFT: short-time Fourier transform)이나 가보 변환(Gabor transform)이 단일주파수 대역내에서만 한정되는 문제를 대체할 새로운 크기의 고정된(fixed) 필터 창 함수(window)를 사용하는 반면, 웨이블릿 변환은 고주파 대역에서는 폭이 좁은 창 함수를, 그리고 저주파 대역에서는 폭이 넓은 창 함수를 가변적으로 사용한다. 따라서 웨이블릿 변환은 상대 대역폭 불변 해석(constant relative bandwidth analysis)이라고도 하며 주파수 대역의 변화 폭이 항상 주파수 값과 비례하는 특징을 갖는다.
- [68] 일실시에에 따른 훈련지표 최적화 장치는 웨이블릿 해석에 기반하여 환경 계측 자료를 구성하는 각 수치에 대해 시계열로 발생하는 변화를 고려한다. 따라서, 훈련지표 최적화 장치는 시간 영역과 주파수 영역을 동시에 분석이 가능하며, 연속 신호와 이산 신호를 모두 모니터링할 수 있다.
- [69] 또한, 일실시에에 따른 훈련지표 최적화 장치는 핵심 특성치 그룹(key features group)을 선정하고, 선정된 핵심 특성치 그룹과 환경 계측 자료를 입력으로 활용하여 환경 계측 자료의 핵심적인 시공간적 변화특성(spatiotemporal feature)을 보다 집중적으로 학습시킴으로써 관측치와 요인에 대한 예측력을 지속적으로 향상시킬 수 있다.
- [70] 즉, 일실시에에 따른 훈련지표 최적화 장치는 핵심 특성치 그룹과 환경 계측 자료를 입력으로 활용하는 심층신경망학습모델(deep learning model)을 구축하여, 환경 계측 자료의 자연배경변동의 핵심적인 시공간적 변화특성을 집중적으로 학습시킴으로써, 관측치와 요인에 대한 예측력을 지속적으로 향상시킬 수 있다.
- [71] 결국, 본 발명을 이용하면, 수문-환경 순환(hydro-environmental cycle)과 연관되어 관측되는 하천수, 지하수의 수위 및 EC (RWL, GWL, RWEC, GWEC), Soil gas concentration and flux (CO₂, FCO₂, CH₄, C₂H₆)의 주요 구동력을 식별하고 기여도를 평가할 수 있으며, 관측자료와 주요환경구동력을 입력자료로 하는 Long-short term memory network (LSTM)을 이용한 deep learning 예측모델을 통해 RWL/RWEC/GWL/GWEC/CO₂/FCO₂/CH₄/C₂H₆에 대한 예측력 향상과 ANN 모델의 최대 난점 중 하나인 hyper parameter 선정의 인과성 및 훈련 가중치 산정의 비선형성 문제에 대한 해석의 근거를 제시할 수 있다.
- [72] 보다 구체적인 예를 들면, 본 발명은 기설정된 기간동안 계측된 일단위 하천수위, 지하수위, 강우자료를 이용하여 지하수위 변동의 주요 구동력을 식별하고 기여도를 평가할 수 있으며, 107일간 계측된 6시간단위 수문기상변수(강우량, 대기온도, 상대습도, 일사량, 풍속), 깊이에 따른 토양특성변수(토양수분, 토양전기전도도, 토양온도) 및 토양호흡변수 (CO₂

농도, CO₂ flux, 수분함량)를 이용하여 지표 CO₂ 플럭스의 주요 구동력을 식별하고 기여도를 정량적으로 평가할 수도 있다.

- [73] 일실시에에 따른 훈련지표 최적화 장치에 관한 구성은 이후 실시예 도 2를 통해 보다 구체적으로 설명하기로 한다.
- [74] 도 2는 일실시에에 따른 훈련지표 최적화 장치의 구성을 설명하기 위한 도면이다.
- [75] 다시 말해, 도 2는 도 1을 통해 설명한 일실시에에 따른 훈련지표 최적화 장치의 구현 예를 설명하는 도면으로, 이후 도 2를 통해 설명하는 내용 중 일실시에에 따른 훈련지표 최적화 장치를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.
- [76] 도 2를 참조하면, 일실시에에 따른 훈련지표 최적화 장치(200)는 데이터의 특성과 모델링 목적에 따른 최적 훈련조건을 사전에 결정할 수 있다.
- [77] 또한, 훈련지표 최적화 장치(200)는 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 훈련시간 단축, 핵심 특성에 대한 집중학습 및 과적합을 방지할 수 있다.
- [78] 또한, 훈련지표 최적화 장치(200)는 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 입력자료에 대한 실시간 모델 업데이트를 통해 예측능을 보다 더 향상시킴과 동시에 최적훈련지표 선정을 위한 정량적 근거를 제시함으로써 다양한 환경자료들을 이용한 통합환경모니터링, 해석, 예측 및 대응시스템 구축에 보다 효율적인 환경 예측 모델을 제공할 수 있다.
- [79] 즉, 일실시에에 따른 훈련 지표 최적화 장치(200)는 기존 LSTM 모델과 관련된 딥 러닝(deep learning) 기법들의 공통적인 한계점인 사전 훈련 최적화 조건을 설정(훈련지표 최적화)하고, 설정 이유와 사후(훈련 후) 예측능에 대한 인과성 판단의 정량적인 근거를 제시할 수 있다.
- [80] 이를 위해, 훈련지표 최적화 장치(200)는 전처리부(210), 동적특성 처리부(220), 핵심 특성치 그룹 선정부(230) 및 지표 최적화부(240)를 포함할 수 있다.
- [81] 일실시에에 따른 전처리부(210)는 환경 계측 자료에 대한 기반데이터세트를 구성할 수 있다.
- [82] 일측에 따르면, 환경 계측 자료는 실시간으로 계측되는 수문-환경 시계열 자료를 포함할 수 있으며, 수문-환경 시계열 자료는 수문 기상 데이터, 하천 수위 데이터, 지하 수위 데이터, 수질 데이터, 온도 데이터, EC 데이터, 동위원소비율 데이터, 토양 가스 데이터 및 미세먼지 데이터 중 적어도 하나 이상의 환경 데이터를 포함할 수 있다.
- [83] 일측에 따르면, 전처리부(210)는 환경 계측 자료의 데이터세트의 관측항목과 관측시간 해상도에 따른 데이터 매트릭스를 기반데이터세트로 구성 및 정렬하고, 정렬된 기반데이터세트에 대한 시간영역 해상도 또는 시간관측 간격별 결측자료를 보간 처리 하고, 보간 처리된 기반데이터세트의 자료를 노이즈 필터링하며, 노이즈 필터링된 결과를 표준화 및 정규화할 수 있다.

- [84] 예를 들면, 전처리부(210)는 복합 환경시계열 계측자료 데이터셋인 기반데이터세트의 관측항목과 관측시간 해상도(스케일)의 재조정 및 목적기간에 따른 변수항목 재조정을 통해 격간 매트릭스(panel matrix)를 구성 및 정렬할 수 있으며, 이렇게 정렬되는 매트릭스가 기반 데이터세트로 해석될 수 있다.
- [85] 예를 들면, 스케일은 시간영역 해상도를 의미할 수 있으며, 시간영역 해상도는 시계열 계측자료의 시간 영역이 기 설정된 주기(period)로 구분된 복수의 영역 각각을 의미할 수 있다.
- [86] 또한, 결측자료는 레코드에 해당하는 데이터가 존재하지 않거나 레코드에 대해 자료가 없는 요소로 해석될 수 있다.
- [87] 또한, 전처리부(210)는 푸리에/웨이블렛 기반(Fourier/wavelet-based)의 다해상도 필터뱅크를 이용하여 노이즈 필터링을 수행할 수 있다.
- [88] 한편, 전처리부(210)는 노이즈 필터링된 결과를 기 공지된 표준화 및 정규화 기법을 통해 표준화 및 정규화할 수 있다.
- [89] 일실시에에 따른 동적 특성 처리부(220)는 다해상도 웨이블렛 해석과 차원 축소 기법을 통해 구성된 기반데이터세트에 대한 동적 특성을 식별 및 추출할 수 있다.
- [90] 일측에 따르면, 동적 특성 처리부(220)는 구성된 기반데이터세트에 대한 시간영역 해상도에 따른 다해상도 웨이블렛 해석을 통해 시간-주파수 영역에 대한 웨이블렛 에너지 분포 데이터를 도출하고, 도출된 웨이블렛 에너지 분포 데이터에 대하여 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군(potential environmental drivers, PEDs)을 선정할 수 있다.
- [91] 예를 들면, 잠재 환경 구동력 후보군은 다수의 구동력 중 실제 관측되고 분석된 자료의 구동력이 될 수 있는 후보들을 의미할 수 있다.
- [92] 보다 구체적인 예를 들면, 잠재 환경 구동력 후보군은 주 대상 관측변수의 변화를 유발하는 환경적 원인으로 해석될 수 있으며, 환경 계측 자료가 토양 가스인 경우에 토양가스의 농도/플럭스 및 이들의 시변특성/시공간특성의 변화를 유발하는 원인으로 해석될 수 있다.
- [93] 다시 말해, 동적 특성 처리부(220)는 전처리부(210)를 통해 수신한 기반데이터세트의 시간영역 해상도에 따른 다해상도 스펙트럴 분석 기반의 시간-주파수 영역의 웨이블렛 에너지 분포를 도출할 수 있다.
- [94] 예를 들면, 차원 축소 기법은 주성분/독립성분 분석(principle/independent component analysis; PCA/ICA), 시계열 요인분석(time series factor analysis, TSFA), 경험적모드분해(empirical mode decomposition; EMD) 및 다해상도상태공간모델(multiresolution state-space model; MRSSM) 중 적어도 하나 이상의 기법을 포함할 수 있다.
- [95] 즉, 동적특성 처리부(220)는 상술한 차원 축소 기법을 통해 최적 잠재 환경 구동력 추출을 위한 후보모델들을 선정하고, 선정된 후보모델들 각각에

대응되는 주요 잠재 환경 구동력 후보군을 선정할 수 있다.

- [96] 일측에 따르면, 동적특성 처리부(220)는 선정된 잠재 환경 구동력 후보군의 시간영역 해상도별 시간변화에 따른 변동특성을 추출하고, 추출된 변동특성에 기초하여 동적 특성을 추출 및 정량화할 수 있다.
- [97] 일실시에에 따른 핵심 특성치 그룹 선정부(230)는 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별 및 평가하고, 평가 결과에 대응하여 핵심 특성치 그룹(key features group)을 선정할 수 있다.
- [98] 예를 들면, 환경 계측 자료에 대한 구동력은 실제 환경 계측 자료의 변동을 지배하는 복합적 환경요인을 나타내는 일반 명사를 의미할 수 있다.
- [99] 일측에 따르면, 핵심 특성치 그룹 선정부(230)는 잠재 환경 구동력 후보군의 잠재 환경 구동력과, 환경 계측 자료 간 다해상도 상관성을 판단하되, 잠재 환경 구동력과 관측자료간 시간지연 및 위상변화를 반영한 상관성 판단을 수행하고, 수행된 상관성 판단 결과에 기초하여, 잠재 환경 구동력과 관측자료간 최대상관스케일을 선정할 수 있다.
- [100] 예를 들면, 최대상관스케일은 잠재구동력과 관측자료간의 다해상도 상관관계 및 교차상관관계를 기반으로 하는 최대상관관계의 시간-주파수 밴드(band)일 수 있으며, 최대상관은 다해상도 상관관계 및 교차상관관계의 분석을 통해 도출되는 복수의 결과값들 중에서 상관도가 가장 높은 결과를 의미할 수 있다.
- [101] 일측에 따르면, 핵심 특성치 그룹 선정부(230)는 잠재 환경 구동력과 관측자료간 웨이블릿 에너지 비율과 선정된 최대상관스케일의 상관관계를 이용하여 구동력을 식별하고, 선정된 최대상관스케일의 결합에너지비율과 차원축소모델의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 평가된 상대적 기여도에 기초하여 핵심 특성치 그룹을 선정할 수 있다.
- [102] 다시 말해, 핵심 특성치 그룹 선정부(230)는 잠재 환경 구동력 후보군을 다해상도 시간-주파수 상관성/ 교차 상관성 진단을 통해 식별(identification)하여 (핵심) 구동력을 선정할 수 있다.
- [103] 예를 들면, 상대적 기여도는 최대상관스케일의 결합에너지비율과 최적 잠재 환경 구동력 선정모델들의 구동력 선정 지표들(factor loading or correlations)간의 선형결합을 통한 유효동적 효율(t , D_{ef})에 기초하여 평가될 수 있으며, 여기서, t 는 최대상관스케일 (cycle), D_{ef} 는 기여도 비율(%)을 의미할 수 있다.
- [104] 바람직하게는, 기여도 비율은 하기 수학적식 5를 통해 도출될 수 있다.
- [105] [수학적식5]
- [106]
$$D_{ef}(\%) = \alpha \times E_c \times 100$$
- [107] 여기서, α 는 동적요인 부하량(dynamic factor loading), E_c 는

최대상관스케일의 결합에너지비율일 수 있다.

- [108] 즉, 평가된 상대적 기여도는 어떤 요인에 의해서 기반데이터세트의 급격한 변화가 발생했는지를 판단하는데 사용될 수 있으며, 식별 및 평가된 구동력에 기초하여 발생한 현상에 대한 최적 대응 시나리오를 제공하도록 지원할 수 있다.
- [109] 다시 말해, 핵심 특성치 그룹 선정부(230)는 유효동적 효율(t, D_{ef}) 기반의 최적 LSTM 훈련지표 선정을 위한 목표변수별 핵심 특성치 그룹을 선정할 수 있다.
- [110] 일측에 따르면, 핵심 특성치 그룹 선정부(230)는 잠재 환경 구동력 후보군 및 핵심 특성치 그룹 중 적어도 하나 이상을 이용하여 훈련되는 사전 조정된 LSTM 네트워크(well-tuned LSTM networks)를 구축하고, 사전 조정된 LSTM 네트워크를 이용하여 잠재 환경 구동력을 검증할 수 있다.
- [111] 예를 들면, 사전 조정된 LSTM 네트워크는 최적 훈련 조건을 선정하기 위한 후보군으로 잠재 환경 구동력 후보군을 포함하여 구성되는 LSTM 네트워크일 수 있다.
- [112] 결국, 일실시에에 따른 훈련지표 최적화 장치(200)를 이용하면 시계열 데이터에 기초한 환경 계측 자료에 대해 시간에 따라 변화하는 다양한 환경요인들의 구조적 특성을 반영할 수 있다. 또한, 다해상도 시간-주파수 영역의 해석방법을 사용하여 관측값 변화의 원인이 되는 핵심적인 환경구동력들(key environmental drivers)과 그 효과들을 보다 체계적으로 분리, 식별할 수 있다.
- [113] 일실시에에 따른 지표 최적화부(240)는 선정된 핵심 특성치 그룹과 환경 계측 자료를 입력으로 수신하여 환경 예측 모델에 대응되는 복수의 훈련지표 값을 제어할 수 있다.
- [114] 예를 들면, 복수의 훈련지표 값은 훈련기간(training period, T), 최소 배치 사이즈(minibatch size, mbs), 은닉층(hidden layers, HL)의 개수 및 최적 반복회수(epochs, E) 중 적어도 하나의 값을 포함할 수 있다. 여기서, 배치 사이즈는 1회의 훈련(1 epoch)에 사용되는 훈련 구간의 크기를 의미할 수 있다.
- [115] 일측에 따르면, 지표 최적화부(240)는 핵심 특성치 그룹과 환경 계측 자료를 입력으로 하는 장단기기억네트워크(Long-short Term Memory Network) 모델을 구축하고, 핵심 특성치 그룹의 시간-주파수 영역에 기초하여 복수의 훈련지표 값을 사전 정량화(pre-quantification)할 수 있다.
- [116] 예를 들면, 구축된 장단기기억네트워크 모델은 식별된 핵심 구동력을 이용하여 최고상관주파수 대역에서 보다 집중된 학습을 통해 구축되고, 보다 개선되며, 선별 및 강화된 네트워크일 수 있다.
- [117] 다시 말해, 지표 최적화부(240)는 최대상관스케일의 최고상관주파수 대역에 기초하여 복수의 훈련지표 값을 사전 정량화할 수 있다.
- [118] 여기서, 최대상관스케일은 단일 주파수로 정해지는 것이 아니라, 주파수 구간(band)으로 정해지므로, 주파수 구간(예를 들면, D1 내지 D3)을 특정하고, 다양한 훈련지표들을 조합하여 구축된 장단기기억네트워크 모델을

갱신(update)할 수 있다.

- [119] 일측에 따르면, 지표 최적화부(240)는 환경 예측 자료로부터 측정되는 관측값과, 장단기기억네트워크 모델로부터 예측되는 예측값의 복합모델검증지표에 기초한 잔차검증 및 잔차의 다해상도해석에 기반하여 적어도 하나 이상의 예측 모델을 선정하고, 선정된 적어도 하나 이상의 예측 모델 중 어느 하나의 예측 모델 또는 둘 이상의 예측 모델이 결합한 결합 예측 모델(combined prediction model)에 기초하여 사전 정량화된 복수의 훈련지표를 정량화할 수 있다.
- [120] 예를 들면, 다해상도해석에 기반하여 선정되는 예측 모델은 시나리오 기반의 환경 대응 시스템에서 다양한 시나리오에 대응되는 환경 예측 모델들을 의미할 수 있다. 즉, 다해상도해석에 기반하여 선정되는 예측 모델은 통합환경 모니터링 및 대응 시스템 구축을 위한 환경 예측 모델일 수 있다.
- [121] 일측에 따르면, 지표 최적화부(240)는 AICc X BIC, RMSE(root mean squared error) X MAPE(maximum absolute percentage error) 및 선형성 대표지수 (R^2 or adjusted R^2) 중 적어도 하나 이상의 복합모델검증지표에 기초한 잔차검증 및 잔차에 대한 다해상도 해석을 수행하여 최적훈련지표기반의 최적 예측모델 선정할 수 있다.
- [122] 보다 구체적인 예를 들면, 지표 최적화부(240)는 예측 모델의 정확도를 판단하기 위한 복합모델검증지표(Combined Index (CI), AICc X BIC or RMSE X MAPE)가 최고 성능(AICc X BIC 및 RMSE X MAPE의 경우는 최소값)을 나타내는 예측 모델들을 선정할 수 있다.
- [123] 즉, 단일 예측 모델이 최고 성능을 나타내는 경우도 있지만, 복수의 예측 모델의 결합 예측 모델이 최고의 성능을 발휘하는 경우도 있으므로, 지표 최적화부(240)는 데이터 특성에 따라 최고성능 예측 모델(단일 또는 결합 예측 모델)을 선정하고, 선정된 모델의 복수의 훈련지표를 최적 훈련지표로 간주할 수 있다.
- [124] 일측에 따르면, 지표 최적화부(240)는 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 이용하는 최적훈련지표모형(예를 들면, 일반화선형함수 모형 또는 회귀모형)을 구성할 수 있다.
- [125] 예를 들면, 최적훈련지표모형은 복수의 훈련지표인 은닉층(hidden layers, HL)의 개수, 반복 회수(epoch, E), 최소 배치 사이즈(minibatch size, mbs), 훈련 기간(training period, T)들 간의 (선형)회귀형 관계를 나타내는 모델일 수 있다.
- [126] 보다 구체적인 예를 들면, 최적훈련지표모형은 길이 1095인 하천수EC (JDEC) 일평균 자료 JDEC의 90% 훈련구간, hidden layer=200에 대한 최적 훈련조건(Op_JDEC)을 $Op_JDEC = X mbs^k + YE + z$ (여기서 mbs는 minibatch size, E는 epochs, k, X, Y, z는 추정된 모수(estimated parameter))와 같은 형태로 표현하는 모델일 수도 있다.
- [127] 일실시에에 따른 최적훈련지표모형을 구성하는 예시는 이후 실시예 도 5a 내지

도 5c를 통해 보다 구체적으로 설명하기로 한다.

- [128] 다시 말해, 지표 최적화부(240)는 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 환경 예측 모델에 적용하여 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 훈련시간 단축, 핵심 특성에 대한 집중학습 및 과적합을 방지할 수 있다.
- [129] 또한, 지표 최적화부(240)는 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 입력자료에 대한 실시간 모델 업데이트를 통해 예측능을 보다 더 향상시킴과 동시에 최적훈련지표 선정을 위한 정량적 근거를 제시함으로써 다양한 환경자료들을 이용한 통합환경모니터링, 해석, 예측 및 대응시스템 구축에 보다 효율적인 환경 예측 모델을 제공할 수 있다.
- [130] 도 3a 내지 도 3h는 일실시에에 따른 훈련지표 최적화 장치에서 토양 가스를 환경 계측 자료로 수신하여 동작하는 예시를 설명하기 위한 도면이다.
- [131] 다시 말해, 도 3a 내지 도 3h는 도 1 내지 도 2를 통해 설명한 일실시에에 따른 훈련지표 최적화 장치의 동작예를 설명하는 도면으로, 이후 도 3a 내지 도 3h를 통해 설명하는 내용 중 일실시에에 따른 훈련지표 최적화 장치를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.
- [132] 도 3a 내지 도 3h를 참조하면, 참조부호 310은 시간 변화에 따른 토양 가스 계측 자료의 변화를 기 설정된 시간 단위로 나타내고, 참조부호 320은 시간 변화에 따른 토양 가스 계측 자료의 변화를 초(second) 단위로 나타낸다.
- [133] 또한, 참조부호 330은 토양 가스 계측 자료의 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 나타내고, 참조부호 340은 토양 가스 계측 자료에 대한 잠재 환경 구동력 후보군을 나타낸다.
- [134] 또한, 참조부호 350은 잠재 환경 구동력 후보군과 토양 가스 계측 자료 사이의 상관성 분석 결과를 나타내고, 참조부호 360은 사전 조정된 LSTM 네트워크(well-tuned LSTM networks)를 이용한 잠재 환경 구동력 후보군들의 예측(prediction) 결과를 나타내며, 참조부호 370은 사전 조정된 LSTM 네트워크를 이용한 토양 가스 계측 자료의 예측 결과를 나타낸다.
- [135] 또한, 참조부호 380은 MR-LSTM 모델을 이용한 토양 가스 계측 자료의 전망(forecast) 결과를 나타낸다.
- [136] 참고로, 참조부호 310 내지 380에서 CH₄는 메탄 가스농도, C₂H₆는 에탄 가스농도, CCO₂는 이산화탄소 농도, FCO₂는 CO₂ 플럭스, T_{-cham}은 챔버 내 대기 온도, P_{-cham}은 챔버 내 기압, RH는 상대 습도, H₂O는 토양표면수분함량, H₂O_{-L}은 챔버내 수분함량, T_{-soil}은 토양온도, SWC는 토양체적함수비, PED1 내지 PED5는 제1 내지 제5 잠재 환경 구동력 후보군을 의미할 수 있다.
- [137] 구체적으로, 참조부호 310 내지 320에 따르면, 시간 변화에 따른 토양 가스 계측 자료의 변화에서는 CH₄, C₂H₆ 및 CCO₂를 비롯한 변수 사이에서 명확한 선형 상관성을 도출하기 어려우며, 선형 상관성을 도출하기 위해서 현장 측정 데이터에 지연된 상호 의존성과, 비선형 응답에 기초한 새로운 접근법이

필요함을 알 수 있다.

- [138] 다시 말해, 단일 스케일(시간영역 해상도) 기반의 선형 상관 관계는 복잡하고 비선형성을 갖는 토양 가스의 동적특성을 나타내는데 효과적이지 못하므로, 일실시에에 따른 훈련지표 최적화 장치와 같이 스케일에 따른 상관성을 기반으로 하는 다중 스케일(multi-scale) 상관성 분석이 수행될 필요가 있다.
- [139] 참조부호 330에 따르면, 일실시에에 따른 훈련지표 최적화 장치는 토양 가스 계측 자료에 대하여 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 도출할 수 있다.
- [140] 보다 구체적으로, 참조부호 330은 토양 가스 계측 자료와 관련한 주요 관측변수의 원시 데이터에 대한 웨이블릿 에너지 분포 데이터를 도출하며, D1 내지 D5와 A1 내지 A5는 이산웨이블릿 분석에 의한 웨이블릿 분해 단계로써, 관측자료의 길이에 따라 다양한 범위에 걸쳐서 웨이블릿 해석에 사용되는 각 성분의 시간-주파수 스케일을 나타낼 수 있다.
- [141] 웨이블릿 에너지 분포를 위한 환경 시계열에서 잠재 환경 구동력 후보군은 DWT(Discrete wavelet transform)를 사용하여 최종 근사성분(A5)과 세부 성분(D1 내지 D5)으로 분해될 수 있다.
- [142] DWT(Discrete wavelet transform)의 수학적식을 구성하는 항목 중에서 분해 레벨을 위한 A_p 와 D_p 가 포함될 수 있는데, A_p 와 D_p 는 각 분해 레벨(p)에서 0.25 사이클 이하의 저주파 신호와 0.25 내지 0.5 사이클의 고주파 신호를 적용할 수 있다. 관측자료의 길이에 따른 분해레벨을 고려할 때 일실시에에 따른 최대 분해 레벨 A5는 32 시간의 스케일에 해당하며, 이는 원자료의 관측간격과 길이에 따라 다양하게 선택될 수 있다.
- [143] 이후, 일실시에에 따른 분해 수준에 대한 모든 시간 주파수 척도는 2시간(D1), 4시간(D2), 8시간(D3), 16시간(D4), 32시간(D5)이 될 수 있다.
- [144] 일례로, D1에서 D3까지의 프로세스는 단기(8 시간의 스케일)로 간주 될 수 있으며 D5와 A5의 프로세스는 장기(32 시간의 스케일 또는 그 이상) 또는 계절성으로 간주 될 수 있다.
- [145] 참조부호 330에서 보는 바와 같이, 일실시에에 따른 훈련지표 최적화 장치는 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터 분석을 통해 복잡한 환경 요인들에 대한 잠재 환경 구동력 후보군(PED1 내지 PED5)의 영향력을 판단할 수 있다.
- [146] 참조부호 340에 따르면, 일실시에에 따른 훈련지표 최적화 장치는 다해상도 웨이블릿 해석을 통해 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 도출하고, 도출된 웨이블릿 에너지 분포 데이터에 대하여 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군을 선정할 수 있다.
- [147] 참조부호 340에 의해 선정된 잠재 환경 구동력 후보군은 하기 표1과 같이 표현될 수 있다.

[148] [표1]

PEDs	Variables with loading	Environmental Meaning
1	CH ₄ (0.59), T _{cham} (0.12)	CH ₄ dynamics with T _{atm} effect
2	P _{cham} (-0.11)	Air movement effect
3	C ₂ H ₆ (0.74), RH(-0.26)	C ₂ H ₆ dynamics with RH effect
4	T _{cham} (1.00), RH (-0.88), CCO ₂ (-0.59), FCO ₂ (-0.25)	CO ₂ dynamics with T _{atm} and RH effect
5	T _{soil} (0.58), SWC (-0.51)	Soil temperature and moisture effect

[149] 참조부호 350에 따르면, 일실시에에 따른 훈련지표 최적화 장치는 잠재 환경 구동력 후보군과 토양 가스 계측 자료 사이의 다해상도 상관성을 판단할 수 있다.

[150] 보다 구체적으로, 일실시에에 따른 훈련지표 최적화 장치는 관측자료들 중에서 어느 주파수가 핵심 주파수인지 여부를 확인하기 위해서는 스케일 대비 상관도를 고려할 수 있다.

[151] 관측자료들 구성하는 각 요소들에 대해, 스케일에 대비하는 상관 계수(r)를 확인하여 어떤 주파수가 핵심 주파수인지 여부를 판단할 수 있다. 예를 들어, 각 이산웨이블렛 분해레벨당 상관계수(r)가 0.5 이상인 경우에 대하여 PED1은 D2의 스케일에서 높은 상관 계수를 보이고, PED3는 D5의 스케일에서 높은 상관계수를 보이며, PED4는 D4의 스케일에서 높은 상관계수를 보인다.

[152] 즉, 일실시에에 따른 훈련지표 최적화 장치는 웨이블렛 에너지 분포를 확인함으로써, 상관도가 가장 높은 최대상관스케일을 선정할 수 있다.

[153] 한편, 일실시에에 따른 훈련지표 최적화 장치는 선정된 최대상관스케일의 결합에너지비율과 적절한 차원축소모델(예를 들면, 다해상도상태공간모델)의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 평가된 상대적 기여도에 기초하여 핵심 특성치 그룹을 선정할 수 있다.

[154] 일측에 따르면, 상대적 기여도는 최대상관스케일의 결합에너지비율과 최적 잠재 환경 구동력 선정모델들의 구동력 선정 지표들(factor loading or correlations)간의 선형결합을 통한 유효동적 효율(t, D_{ef})에 기초하여 평가될 수 있으며, 평가된 상대적 기여도는 하기 표 2와 같이 나타낼 수 있다.

[155] [표2]

Time scales	Soil gases	Potential Environmental drivers (PEDs)	Main bands (Scales)	Cumulative energy ratio (Ec)	DyF loading(α)	Maximum Def (%)
DW_SGs_1h	CH ₄	Complex effect of CH ₄ emission & T _{atm} (PED1)	D1-D2 (2-4h)	0.28	0.59	16.52%
	C ₂ H ₆	RH effect (PED3)	D5-A5 (32h)	0.13	0.74	9.62%
	CCO ₂	Complex effect of CO ₂ emission and T _{atm} +RH (PED4)	D4 (16h)	0.54	-0.59	31.86%
	FCO ₂	Seasonality	D5-A5 (32h)	0.33	-0.25	8.25%

- [156] 표 2에 따르면, 상대적 기여도는 PED4 및 PED1이 가장 높은 것으로 평가되었으며, 일실시에에 따른 훈련지표 최적화 장치는 평가된 상대적 기여도에 기초하여 핵심 특성치 그룹을 선정할 수 있다.
- [157] 참조부호 360 내지 380에 따르면, 일실시에에 따른 훈련지표 최적화 장치의 예측(prediction) 및 전망(forecast) 능력은 선정된 토양 가스의 핵심 특성에 기초한 강화된 집중학습(focused learning)을 통해 개선될 수 있음을 확인할 수 있다.
- [158] 도 4a 내지 도 4f는 일실시에에 따른 훈련지표 최적화 장치에서 하천/ 지하수 수위(RWL/GWL) 및 전기 전도도(RWEC/ GWEC) 계측자료를 환경 계측 자료로 수신하여 동작하는 예시를 설명하기 위한 도면이다.
- [159] 다시 말해, 도 4a 내지 도 4f는 도 1 내지 도 3h를 통해 설명한 일실시에에 따른 훈련지표 최적화 장치의 동작예를 설명하는 도면으로, 이후 도 4a 내지 도 4f를 통해 설명하는 내용 중 일실시에에 따른 훈련지표 최적화 장치를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.
- [160] 도 4a 내지 도 4f를 참조하면, 참조부호 410은 지하수 수위(GWL)/ 전기전도도(EC)에 대한 웨이블릿 에너지 분포 데이터를 나타내고, 참조부호 420은 다해상도상대공간모델(MRSSM)을 이용하여 지하수-전기전도도(GWEC) 계측 자료의 잠재 환경 구동력 후보군을 선정하는 예시를 나타낸다.
- [161] 또한, 참조부호 430은 잠재 환경 구동력 후보군과 지하수-전기전도도(GWEC) 계측 자료 사이의 상관성 분석 결과를 나타내고, 참조부호 440은 사전 조정된 LSTM 네트워크(well-tuned LSTM networks)를 이용한 강물의 전기전도도(JDEC)의 예측(prediction) 결과를 나타낸다.
- [162] 또한, 참조부호 450은 사전 조정된 MR-LSTM 네트워크(well-tuned MR-LSTM networks)를 이용한 GWEC의 예측 결과를 나타내고, 참조부호 460은 사전 조정된

MR-LSTM 네트워크를 이용한 GWEC의 전망(forecast) 결과를 나타낸다.

- [163] 참고로, 참조부호 410 내지 460에서 EC는 지하수/하천수(강물) 전기 전도도, GW는 지하수, JD는 강물 수위 (RWL) 관측소 명, JDEC는 강물의 전기전도도(RWEC), Rainfall은 누적 강우량, PED1 내지 PED4는 전기전도도 (EC)의 제1 내지 제2 잠재 환경 구동력 후보군, PHD1 내지 PHD5는 지하수위(GWL)의 제1 내지 제5 잠재환경 구동력 후보군을 의미할 수 있다.
- [164] 구체적으로, 참조부호 410에 따르면, GWL과 GWEC는 동일한 지하수관측정(GW wells) 에서도 서로 다른 패턴을 가지고 있는 것을 알 수 있으며, 서로 다른 패턴은 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 통해 쉽게 식별 및 정량화가 가능하다는 것을 확인할 수 있다.
- [165] 구체적으로, 참조부호 410의 좌측 하단 도면은 하천수위(강물수위) 또는 하천(강물)EC, 및 지하수위/EC를 나타내고, 참조부호 410의 좌측 상단 도면은 웨이블릿 에너지 밀도 그래프를 나타내며, 좌측 하단에서 관측되는 강과 인접하는 지하수관정(HAM004)의 지하수위는 강물의 수위(RWL=JD)변동과 거의 같게 관측되나 EC의 경우 매우 다르게 관측됨을 확인할 수 있다(좌측 하단 그림 RWEC & GWEC 참조).
- [166] 따라서 이를 웨이블릿에너지 분포로 살펴보면 서로 유사하게 보이는 지하수위 (HAM004) & 하천수위 (JD)도 보다 명확하게 구분할 수 있고, 무엇보다도 수위와 EC의 패턴 차이를 보다 극명하게 표현할 수 있다.
- [167] 이를 기반으로 어느 주파수 영역이 각 수위와 EC의 변동성을 잘 대표하는지를 보다 정량적으로 식별 할 수 있으므로, 잠재 구동력 후보군의 웨이블릿 에너지 분포와 비교/대비 및 상관성/교차상관성 분석을 통해 최고상관 주파수대역을 선정 할 수 있다.
- [168] 예를 들면, RWL(JD)의 경우 일-월 변동성이 30% 정도를 차지함에 반해 RWEC는 15% 미만이며(동일한 관측지점에 대한 수위와 EC의 비교), HAM004EC에 비해 HAM062EC는 장주기 순환변동이 지배하고 있음을 확인할 수 있다(동일한 관측값(EC)에 대한 서로 다른 지점의 변동특성 비교).
- [169] 또한, RWEC(JDEC) 는 GWEC에 비해 전반적으로 단주기 변동성이 더욱 지배적인 것을 확인할 수 있다(동일한 관측 자료(EC) 의 서로 다른 system (강물 & 지하수) 에서의 변동성 비교).
- [170] 결국, 참조부호 410에 따르면, 서로 유사하거나 차이가 많이 나는 복합적 패턴들을 보다 분명하고, 정량적으로 규명할 수 있음을 확인할 수 있다.
- [171] 또한, 참조부호 420에 따르면, 일실시에에 따른 훈련지표 최적화 장치는 GWEC 계측 자료에 대하여 다해상도 웨이블릿 해석을 통해 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 도출하고, 도출된 웨이블릿 에너지 분포 데이터에 대하여 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군을 선정할 수 있다.
- [172] 또한, 참조부호 430에 따르면, 일실시에에 따른 훈련지표 최적화 장치는 잠재

환경 구동력 후보군과 GWEC 계측 자료 사이의 다해상도 상관성을 판단할 수 있다.

- [173] 보다 구체적으로, 일실시에에 따른 훈련지표 최적화 장치는 관측자료들 중에서 어느 주파수가 핵심 주파수인지 여부를 확인하기 위해서는 스케일 대비 상관도를 고려할 수 있다.
- [174] 관측자료를 구성하는 각 요소들에 대해, 스케일에 대비하는 상관 계수(r)를 확인하여 어떤 주파수가 핵심 주파수인지 여부를 판단할 수 있다. 예를 들어, 각 이산웨이블렛 분해레벨당 상관계수(r)가 0.5 이상인 경우에 대하여 PED1의 HAM062EC는 D6의 스케일에서 가장 높은 상관계수를 보이고, HAM060EC는 A6에서 가장 높은 상관계수를 보인다.
- [175] 즉, 일실시에에 따른 훈련지표 최적화 장치는 웨이블렛 에너지 분포를 확인함으로써, 상관도가 가장 높은 최대상관스케일을 선정할 수 있다.
- [176] 참조부호 440 내지 460에 따르면, 일실시에에 따른 훈련지표 최적화 장치의 예측(prediction) 및 전망(forecast) 능력은 선정된 GWEC의 핵심 특성에 기초한 강화된 집중학습(focused learning)을 통해 개선될 수 있음을 확인할 수 있다.
- [177] 구체적으로, 참조부호 450의 좌측 도면은 raw-trained LSTM(관측자료만을 이용하여 훈련된 LSTM)의 출력을 나타내고, 참조부호 460의 우측 도면은 PED trained LSTM(잠재구동력을 통해 훈련된 LSTM)의 출력을 나타낸다.
- [178] 참조부호 450에 따르면, PED를 통해 훈련된 모델이 특히 비선형 자료(예를 들면, HAM047EC와 같이 하강추세가 있는 경우, HAM013EC와 같이 급격한 등락과 상향추세가 있는 경우)의 예측에서 관측자료만을 이용하여 훈련된 모델보다 뛰어난 예측(prediction) 성능을 발휘함을 확인할 수 있다.
- [179] 참조부호 460의 좌측 도면은 비선형 자료(HAM007EC)에 대한 관측자료만으로 훈련된 LSTM (raw LSTM)의 미래 전망 (forecast)의 예시를 나타내고, 참조부호 460의 우측 도면은 비선형 자료(HAM007EC)에 대한 잠재구동력 기반으로 훈련된 LSTM (PED-LSTM)의 미래 전망 예시를 나타낸다.
- [180] 참조부호 460에 따르면, 참조부호 460의 좌측 도면에서 HAM007EC와 같이 급격한 변동 패턴(transient variation)을 보이는 자료의 경우 관측치만으로 훈련된 모델(raw-LSTM)의 미래예측(전망, forecast) 값은 매우 불안정하여 신뢰하기 어려운 경우가 다수 존재한다는 것을 확인할 수 있다.
- [181] 반면, 참조부호 460의 우측 도면에서 PED를 통해 훈련된 모델(PED-LSTM)은 특히 비선형, 과도 변동이 포함된 관측치의 전망에서 raw-LSTM에 비해 보다 안정적이고 납득할 만한 미래예측값을 나타나는 것을 확인할 수 있다.
- [182] 보다 구체적으로, 지하수위, EC, 토양가스농도 및 flux등과 같은 현장규모로 관측되는 환경자료들은 다양한 주기의 수문순환과 인위적 요인들의 복합적인 영향으로 인해 비선형적, 과도적 계측값들을 포함하며, 이들 비선형/과도 계측값은 기 가해진 환경 충격(environmental impacts)에 대한 해당 환경계(environmental system)의 동적반응(dynamic response)를 나타내므로, 정상성

가정을 만족시키기 위해 아웃라이어(outlier)로 처리되거나 제거 시키는 경우에 가장 중요한 동적 반응정보를 유실할 수 있다.

- [183] 이에 PED를 통해 훈련된 LSTM은 이러한 비선형적, 과도신호들의 특성과 예측모델의 안정성 또한 적절히 조화시킬 수 있는 한 대안이 될 수 있으며, 특히 raw-LSTM과 PED-LSTM을 병용하여 다양한 미래시나리오 기반의 실시간 통합환경모니터링 및 대응 시스템을 구축할 수 있다.
- [184] 예를 들면, 해당 GWEC가 2번 모델(LSTM2)을 따른다고 가정하는 경우(2 번 시나리오) raw-LSTM2(좌)과 PED2-LSTM2(우, PED2로 훈련된 LSTM의 2번째 시나리오)의 미래전망치 비교를 통해 Threshold value(예를 들면 2σ 범위 기준)를 선정하거나, 기 설정된 threshold value 대비 조기경보 (early alarm)를 발효 할 수 있다(4f 좌측 raw-LSTM2기준, 1-July-2015).
- [185] 반면 우측의 PED2-LSTM2의 시나리오에 기반하는 경우 좌측 시기의 조기경보의 경우 false alarm으로 판단 할 수도 있으며, 실제 alarm은 1-September-2016 (4f 우측 PED2-LSTM2 기준)에 발효 하는 것으로 상정해 볼 수도 있다.
- [186] 도 5a 내지 도 5c는 일실시에에 따른 최적훈련지표모형을 구성하는 예시를 설명하기 위한 도면이다.
- [187] 다시 말해, 도 5a 내지 도 5c는 도 1 내지 도 4f를 통해 설명한 일실시에에 따른 훈련지표 최적화 장치에서 최적훈련지표모형을 구성하는 예시를 설명하는 도면으로, 이후 도 5a 내지 도 5c를 통해 설명하는 내용 중 일실시에에 따른 훈련지표 최적화 장치를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.
- [188] 도 5a 내지 도 5c를 참조하면, 참조부호 510 내지 530은 하천수 EC (RWECE=JDEC) 자료의 분석을 통해 산출되는 최소 배치 사이즈(minibatch size, mbs) 및/또는 최적 반복회수(epochs, E)를 이용하여 최적훈련지표모형을 구성하는 예시를 설명하기 위한 도면이다.
- [189] 구체적으로, 참조부호 510은 하천수 EC(JDEC) 데이터에 대한 RMSE(root mean squared error)의 경험 법칙(rules of thumb) 분석에 기초하여 최적 mbs(best mini-batch size) 모델을 선택하는 예시를 설명하기 위한 도면이고, 참조부호 520은 하천수 EC(JDEC) 데이터에 대한 상위 6개의 mbs 모델을 통해 최적훈련지표모형을 결정하는 예시를 설명하기 위한 도면이다.
- [190] 또한, 참조부호 530은 하천수 EC (JDEC) 데이터에 대한 RMSE의 경험 법칙 분석에 기초하여 최적 epochs 모델을 선택하는 예시를 설명하기 위한 도면이다.
- [191] 한편, 참조부호 510 내지 530에서 점선형태의 회귀 직선(regression line) 및 번호들(① 내지 ③, a1 내지 a3)은 RMSE 변화(추세)의 주요변곡점(elbow point)을 확인하기 위한 경험법칙(rules of thumb) 적용사례를 보다 구체적으로 설명하기 위한 것이다.
- [192] 구체적으로, 참조부호 510의 첫번째 그림은 1 내지 360 (1, 2, 3, ..., 360)까지의 각 mbs에 대한 RMSE의 변화(HL=200, E=1500)를 나타내고, 두번째 그림은

RMSE 기반의 순위별(작은값 우선) 해당 mbs(1 내지 360)에 따른 RMSE변화를 나타낸다. 여기서, 주요 RMSE 변곡점은 번호 ①에서 상위 5%, 번호 ②에서 상위 5% 내지 60%, 번호 ③에서 상위 60% 내지 100%로 구분될 수 있다.

[193] 또한, 참조부호 510의 세번째 그림은 상위 5%를 다시 주요 RMSE변곡점을 기준으로 최적 mbs 구간인 ①-①(Rank1~6)과, ①-②(Rank 7~18)로 재구분하고, Rank1~Rank6 에 해당하는 모델을 최적 mbs 모델로 선정하는 예시를 나타낸다.

[194] 한편, 참조부호 530의 첫 번째 그림은 1 내지 120까지의 각 E(epochs: 100, 200, 300, ... , 12000)에 대한 RMSE 변화와 최적 RMSE 회귀 직선들의 변곡점(a1 내지 a3), 최적 RMSE (즉, 최소 RMSE)가 나타나는 epoch (E) 구간(참조부호 530의 점선 박스)을 나타내고, 두 번째 그림은 RMSE 기반의 순위별(작은 값 우선) 해당 epoch (E) (100 내지 12000)에 따른 RMSE변화를 나타내며(여기서, 분홍색 박스는 상위 5%), 세번째 그림은 상위 5% (①-①) 와 상위 5~15% (①-②) 까지의 RMSE 순위에 따른 각 epoch (E)별 RMSE regression line의 주요 변곡점 및 최적 epoch (E) 선정 방법의 예시(rules of thumb)를 나타낸다.

[195] 참조부호 510 내지 520에 따르면, 참조부호 510에 도시된 하천수 EC(JDEC) 데이터의 분석 결과를 하기 표3과 같이 표현할 수 있고, 참조부호 520에 도시된 상위 6개의 mbs 모델의 분석 결과를 하기 표4와 같이 표현할 수 있다.

[196] [표3]

Rank	mbs (days)	Cycle (months)	RMSE
1	94	3.1	18.02
2	279	9.3	18.36
3	277	9.2	18.98
4	296	9.9	19.20
5	39	1.3	19.30
6	325	10.8	19.69
7	146	4.9	19.77
8	199	6.6	19.78
9	187	6.2	19.82
10	5	0.2	19.88
11	298	9.9	19.90
12	141	4.7	19.91
13	139	4.6	19.92
14	117	3.9	19.99
15	356	11.9	20.21
16	107	3.6	20.43
17	342	11.4	20.49
18	70	2.3	20.51

[197] [표4]

LSTM_rank	LSTM1	LSTM2	LSTM3	LSTM4	LSTM5	LSTM6	LSTM_avg
mbs (days)	94	279	277	296	39	325	218.3
RMSE	18.02	18.36	18.98	19.20	19.30	19.69	15.74
MAPE(%)	4.4%	4.9%	4.5%	4.6%	4.8%	4.8%	4.0%
Combined Index (CI, RMSE×MAPE)	0.793	0.903	0.846	0.883	0.936	0.953	0.622
Rank_by_CI	1	4	2	3	5	6	-

- [198] 표 3 내지 표 4에 따르면, 일실시에에 따른 훈련지표 최적화 장치는 참조부호 510에 도시된 하천수 EC(JDEC) 데이터에서 RMSE의 경험 법칙 분석에 기초하여 최상위 5%로서 상위 6개의 mbs 모델(표1의 Rank 1 내지 6)을 선정할 수 있으며, 이때 HL (hidden layers)은 200, Epochs (E)는 1500, 훈련기간 (T)은 953일, mbs는 1 내지 360까지 1구간씩 순차적으로 진행되는 범위일 수 있다. 여기서, HL(hidden layers)은 은닉층의 개수를 의미할 수 있다.
- [199] 또한, 훈련지표 최적화 장치는 JDEC와 PEDs 간의 웨이블릿에너지(Ew) 분포 재진단을 통한 최대상관스케일(Ec)과의 상관성 분석 수행할 수 있으며, 이를 통해 최적 mbs 모델 선정의 근거를 제시할 수 있다.
- [200] 일측에 따르면, 훈련지표 최적화 장치는 재훈련 및 재적합을 통해 최적 mbs 모델을 산정할 수 있다. 예를 들면, 훈련지표 최적화 장치는 270 내지 300 mbs 범위에 대한 재훈련 및 재적합을 통해 최적 mbs 모델을 산정할 수 있다.
- [201] 한편, 훈련지표 최적화 장치는 상위 6개의 mbs 모델의 예측값(prediction)에 대한 복합모델설명력 지표(Combined Index (CI), RMSEXMAPE) 기반의 최적 mbs 훈련조건을 이용하여, 최적훈련지표모형을 구성할 수 있다.
- [202] 참조부호 530에 따르면, 참조부호 530에 도시된 하천수 EC(JDEC) 데이터의 분석 결과를 하기 표5과 같이 표현할 수 있다.

[203] [표5]

Rank	mbs (days)	cycle (months)	epochs	RMSE
1	256	8.5	7800	17.64
2	256	8.5	6200	17.90
3	256	8.5	7900	18.03
4	256	8.5	4200	18.25
5	256	8.5	9500	18.41
6	256	8.5	5800	18.61
7	256	8.5	5100	18.65
8	256	8.5	8600	18.70
9	256	8.5	6700	18.74
10	256	8.5	4700	18.90
11	256	8.5	10100	19.14
12	256	8.5	7200	19.26
13	256	8.5	11900	19.35
14	256	8.5	9400	19.53
15	256	8.5	9800	19.55
16	256	8.5	4400	19.57
17	256	8.5	6300	19.77
18	256	8.5	6500	19.98

[204] 표 5에 따르면, 훈련지표 최적화 장치는 참조부호 530에 도시된 하천수 EC(JDEC) 데이터에서 RMSE의 경험 법칙 분석에 기초하여 최상위 5%인 6개의 mbs 모델(표3의 Rank 1 내지 6)을 선정할 수 있으며, 이때 HL (hidden layers)은 200, mbs는 256, 훈련기간은 953일, epochs(E)는 100 내지 12000일 수 있다. 여기서, epochs(E)는 100 부터 100씩 증가하여 12000까지 도달할 수 있다.

[205] 또한, 훈련지표 최적화 장치는 동일한 mbs 구간에 대한 최적 epochs(E) 선정의 근거를 제시할 수 있다.

[206] 일측에 따르면, 훈련지표 최적화 장치는 재훈련 및 재적합을 통해 최적 epochs(E) 모델을 산정할 수 있다. 예를 들면, 훈련지표 최적화 장치는 5000 내지 8000 epochs(E) 범위에 대한 재훈련 및 재적합을 통해 최적 epochs(E) 모델을 산정할 수 있다.

[207] 한편, 훈련지표 최적화 장치는 최상위 6개의 epochs(E) 모델의 예측값(prediction)에 대한 복합모델설명력 지표(CI, RMSEXMAPE) 기반의 최적 epochs(E)의 범위를 보다 구체적으로 설정할 수 있으며, 동시에 이를 이용하여 최적훈련지표모형을 구성할 수 있다.

[208] 도 6은 일실시에에 따른 훈련지표 최적화 방법을 설명하기 위한 도면이다.

[209] 다시 말해, 도 6은 도 1 내지 도 5c를 통해 설명한 일실시에에 따른 훈련지표 최적화 장치의 동작 방법을 설명하는 도면으로, 이후 도 6을 통해 설명하는 내용 중 일실시에에 따른 훈련지표 최적화 방법을 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.

[210] 도 6을 참조하면, 610 단계에서 일실시에에 따른 훈련지표 최적화 방법은

- 전처리부에서 환경 계측 자료에 대한 기반데이터세트를 구성할 수 있다.
- [211] 다음으로, 620 단계에서 일실시에에 따른 훈련지표 최적화 방법은 동적 특성 처리부에서 다해상도 웨이블렛 해석과 차원 축소 기법을 통해 구성된 기반데이터세트에 대한 동적 특성을 식별 및 추출할 수 있다.
- [212] 일측에 따르면, 620 단계에서 일실시에에 따른 훈련지표 최적화 방법은 구성된 기반데이터세트에 대한 시간영역 해상도에 따른 다해상도 웨이블렛 해석을 통해 시간-주파수 영역에 대한 웨이블렛 에너지 분포 데이터를 도출하고, 도출된 웨이블렛 에너지 분포 데이터에 대하여 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군을 선정하는 단계를 더 포함할 수 있다.
- [213] 또한, 620 단계에서 일실시에에 따른 훈련지표 최적화 방법은 선정된 잠재 환경 구동력 후보군의 시간영역 해상도별 시간변화에 따른 변동특성을 추출하고, 추출된 변동특성에 기초하여 동적 특성을 추출 및 정량화하는 단계를 더 포함할 수 있다.
- [214] 다음으로, 630 단계에서 일실시에에 따른 훈련지표 최적화 방법은 핵심 특성치 그룹 선정부에서 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별 및 평가하고, 평가 결과에 대응하여 핵심 특성치 그룹(key features group)을 선정할 수 있다.
- [215] 일측에 따르면, 630 단계에서 일실시에에 따른 훈련지표 최적화 방법은 잠재 환경 구동력 후보군의 잠재 환경 구동력과, 환경 계측 자료 간 다해상도 상관성을 판단하되, 잠재 환경 구동력과 관측자료간 시간지연 및 위상변화를 반영한 상관성 판단을 수행하고, 수행된 상관성 판단 결과에 기초하여, 잠재 환경 구동력과 관측자료간 최대상관스케일을 선정하는 단계를 더 포함할 수 있다.
- [216] 또한, 630 단계에서 일실시에에 따른 훈련지표 최적화 방법은 잠재 환경 구동력과 관측자료간 웨이블렛 에너지 비율과 선정된 최대상관스케일의 상관관계를 이용한 구동력을 식별하고, 선정된 최대상관스케일의 결합에너지비율과 차원축소모델의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 평가된 상대적 기여도에 기초하여 핵심 특성치 그룹을 선정하는 단계를 더 포함할 수 있다.
- [217] 다음으로, 640 단계에서 일실시에에 따른 훈련지표 최적화 방법은 지표 최적화부에서 선정된 핵심 특성치 그룹과 환경 계측 자료를 입력으로 수신하여 환경 예측 모델의 복수의 훈련지표 값을 제어할 수 있다.
- [218] 예를 들면, 복수의 훈련지표 값은 훈련기간(training period, T), 최소 배치 사이즈(minibatch size, mbs), 은닉층(hidden layers, HL)의 개수 및 최적 반복회수(epochs, E) 중 적어도 하나의 값을 포함할 수 있다.
- [219] 일측에 따르면, 640 단계에서 일실시에에 따른 훈련지표 최적화 방법은 핵심 특성치 그룹과 환경 계측 자료를 입력으로 하는 장단기기억네트워크(Long-short Term Memory Network) 모델을 구축하고, 핵심 특성치 그룹의 시간-주파수

영역에 기초하여 복수의 훈련지표 값을 사전 정량화(pre-quantification)하는 단계를 더 포함할 수 있다.

- [220] 또한, 640 단계에서 일실시에에 따른 훈련지표 최적화 방법은 환경 계측 자료로부터 측정되는 관측값과, 장단기기억네트워크 모델로부터 예측되는 예측값의 복합모델검증지표에 기초한 잔차검증 및 잔차의 다해상도해석에 기반하여 적어도 하나 이상의 예측 모델을 선정하고, 선정된 적어도 하나 이상의 예측 모델 중 어느 하나의 예측 모델 또는 둘 이상의 예측 모델이 결합한 결합 예측 모델(combined prediction model)에 기초하여 사전 정량화된 복수의 훈련지표를 사후 정량화(post quantification)하는 단계를 더 포함할 수 있다.
- [221] 또한, 640 단계에서 일실시에에 따른 훈련지표 최적화 방법은 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 이용하는 최적훈련지표모형(예를 들면, 일반화선형함수 모형 또는 회귀모형)을 구성하는 단계를 더 포함할 수 있다.
- [222] 결국, 본 발명을 이용하면, 데이터의 특성과 모델링 목적에 따른 최적 훈련조건을 사전에 결정할 수 있다.
- [223] 또한, 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 훈련시간 단축, 핵심 특성에 대한 집중학습 및 과적합을 방지할 수 있다.
- [224] 또한, 최적 훈련조건을 사전에 결정함으로써, 일반적인 LSTM 모델들에 대해서도 입력자료에 대한 실시간 모델 업데이트를 통해 예측능을 보다 더 향상시킴과 동시에 최적훈련지표 선정을 위한 정량적 근거를 제시함으로써 다양한 환경자료들을 이용한 통합환경모니터링, 해석, 예측 및 대응시스템 구축에 보다 효율적인 환경 예측 모델을 제공할 수 있다.
- [225] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬

프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

- [226] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [227] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

청구범위

- [청구항 1] 환경 계측 자료에 대한 기반데이터세트를 구성하는 전처리부;
다해상도 웨이블릿 해석과 차원 축소 기법을 통해 상기 구성된
기반데이터세트에 대한 동적 특성을 식별 및 추출하는 동적 특성 처리부;
상기 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별
및 평가하고, 상기 평가 결과에 대응하여 핵심 특성치 그룹(key features
group)을 선정하는 핵심 특성치 그룹 선정부 및
상기 선정된 핵심 특성치 그룹과 상기 환경 계측 자료를 입력으로
수신하여 환경 예측 모델에 대응되는 복수의 훈련지표 값을 제어하는
지표 최적화부
를 포함하는 훈련지표 최적화 장치.
- [청구항 2] 제1항에 있어서,
상기 환경 계측 자료는,
실시간으로 계측되는 수문-환경 시계열 자료를 포함하고,
상기 수문-환경 시계열 자료는,
수문 기상 데이터, 하천 수위 데이터, 지하 수위 데이터, 수질 데이터, 온도
데이터, EC 데이터, 동위원소비율 데이터, 토양 가스 데이터 및 미세먼지
데이터 중 적어도 하나 이상의 환경 데이터를 포함하는
훈련지표 최적화 장치.
- [청구항 3] 제1항에 있어서,
상기 전처리부는,
상기 환경 계측 자료의 데이터세트의 관측항목과 관측시간 해상도에
따른 데이터 매트릭스를 상기 기반데이터세트로 구성하고 정렬하고,
상기 정렬된 기반데이터세트에 대한 시간영역 해상도 또는 시간관측
간격 별 결측자료를 보간 처리 하고, 상기 보간 처리된 기반데이터세트의
자료를 노이즈 필터링하며, 상기 노이즈 필터링된 결과를 표준화 및
정규화 하는
훈련지표 최적화 장치.
- [청구항 4] 제1항에 있어서,
상기 동적 특성 처리부는,
상기 구성된 기반데이터세트에 대한 시간영역 해상도에 따른 상기
다해상도 웨이블릿 해석을 통해 시간-주파수 영역에 대한 웨이블릿
에너지 분포 데이터를 도출하고, 상기 도출된 웨이블릿 에너지 분포
데이터에 대하여 상기 차원 축소 기법을 적용하여 잠재 환경 구동력
후보군을 선정하는
훈련지표 최적화 장치.
- [청구항 5] 제4항에 있어서,

상기 동적 특성 처리부는,
 상기 선정된 잠재 환경 구동력 후보군의 시간영역 해상도별 시간변화에 따른 변동특성을 추출하고, 상기 추출된 변동특성에 기초하여 상기 동적 특성을 추출 및 정량화하는
 훈련지표 최적화 장치.

[청구항 6] 제1항에 있어서,
 상기 차원 축소 기법은,
 주성분/독립성분 분석(principle/independent component analysis; PCA/ICA),
 시계열 요인분석(time series factor analysis, TSFA),
 경험적모드분해(empirical mode decomposition; EMD) 및
 다해상도상태공간모델(multiresolution state-space model; MRSSM) 중 적어도 하나 이상의 기법을 포함하는
 훈련지표 최적화 장치.

[청구항 7] 제4항에 있어서,
 상기 핵심 특성치 그룹 선정부는,
 상기 잠재 환경 구동력 후보군의 잠재 환경 구동력과, 상기 환경 계측 자료 간 다해상도 상관성을 판단하되, 상기 잠재 환경 구동력과 상기 관측자료간 시간지연 및 위상변화를 반영한 상관성 판단을 수행하고,
 상기 수행된 상관성 판단 결과에 기초하여, 상기 잠재 환경 구동력과 상기 관측자료간 최대상관스케일을 선정하는
 훈련지표 최적화 장치.

[청구항 8] 제7항에 있어서,
 상기 핵심 특성치 그룹 선정부는,
 상기 잠재 환경 구동력과 상기 관측자료간 웨이블릿 에너지 비율과 상기 선정된 최대상관스케일의 상관관계를 이용한 구동력을 식별하고, 상기 선정된 최대상관스케일의 결합에너지비율과 차원축소모델의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 상기 평가된 상대적 기여도에 기초하여 상기 핵심 특성치 그룹을 선정하는
 훈련지표 최적화 장치.

[청구항 9] 제7항에 있어서,
 상기 핵심 특성치 그룹 선정부는,
 상기 잠재 환경 구동력 후보군 및 상기 핵심 특성치 그룹 중 적어도 하나 이상을 이용하여 훈련되는 사전 조정된 LSTM 네트워크(well-tuned LSTM networks)를 구축하고, 상기 사전 조정된 LSTM 네트워크를 이용하여 상기 잠재 환경 구동력을 검증하는
 훈련지표 최적화 장치.

[청구항 10] 제1항에 있어서,
 상기 지표 최적화부는

상기 핵심 특성치 그룹과 상기 환경 계측 자료를 입력으로 하는 장단기기억네트워크(Long-short Term Memory Network) 모델을 구축하고, 상기 핵심 특성치 그룹의 시간-주파수 영역에 기초하여 상기 복수의 훈련지표 값을 사전 정량화(pre-quantification)하는 훈련지표 최적화 장치.

[청구항 11] 제10항에 있어서,
상기 지표 최적화부는,
상기 환경 계측 자료로부터 측정되는 관측값과, 상기 장단기기억네트워크 모델로부터 예측되는 예측값의 복합모델검증지표에 기초한 잔차검증 및 잔차의 다해상도해석에 기반하여 적어도 하나 이상의 예측 모델을 선정하고, 상기 선정된 적어도 하나 이상의 예측 모델 중 어느 하나의 예측 모델 또는 둘 이상의 예측 모델이 결합한 결합 예측 모델(combined prediction model)에 기초하여 상기 사전 정량화된 복수의 훈련지표를 정량화하는 훈련지표 최적화 장치.

[청구항 12] 제11항에 있어서,
상기 지표 최적화부는,
상기 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 이용하는 최적훈련지표모형을 구성하는 훈련지표 최적화 장치.

[청구항 13] 제1항에 있어서,
상기 복수의 훈련지표 값은,
훈련기간(training period, T), 최소 배치 사이즈(minibatch size, mbs), 은닉층(hidden layers, HL)의 개수 및 최적 반복회수(epochs, E) 중 적어도 하나의 값을 포함하는 훈련지표 최적화 장치.

[청구항 14] 전처리부에서, 환경 계측 자료에 대한 기반데이터세트를 구성하는 단계;
동적 특성 처리부에서, 다해상도 웨이블렛 해석과 차원 축소 기법을 통해 상기 구성된 기반데이터세트에 대한 동적 특성을 식별 및 추출하는 단계;
핵심 특성치 그룹 선정부에서, 상기 추출된 동적 특성에 기초하여 환경 계측 자료에 대한 구동력을 식별 및 평가하고, 상기 평가 결과에 대응하여 핵심 특성치 그룹(key features group)을 선정하는 단계 및
지표 최적화부에서, 상기 선정된 핵심 특성치 그룹과 상기 환경 계측 자료를 입력으로 수신하여 환경 예측 모델의 복수의 훈련지표 값을 제어하는 단계
를 포함하는 훈련지표 최적화 방법.

[청구항 15] 제14항에 있어서,
상기 동적 특성을 식별 및 추출하는 단계는,

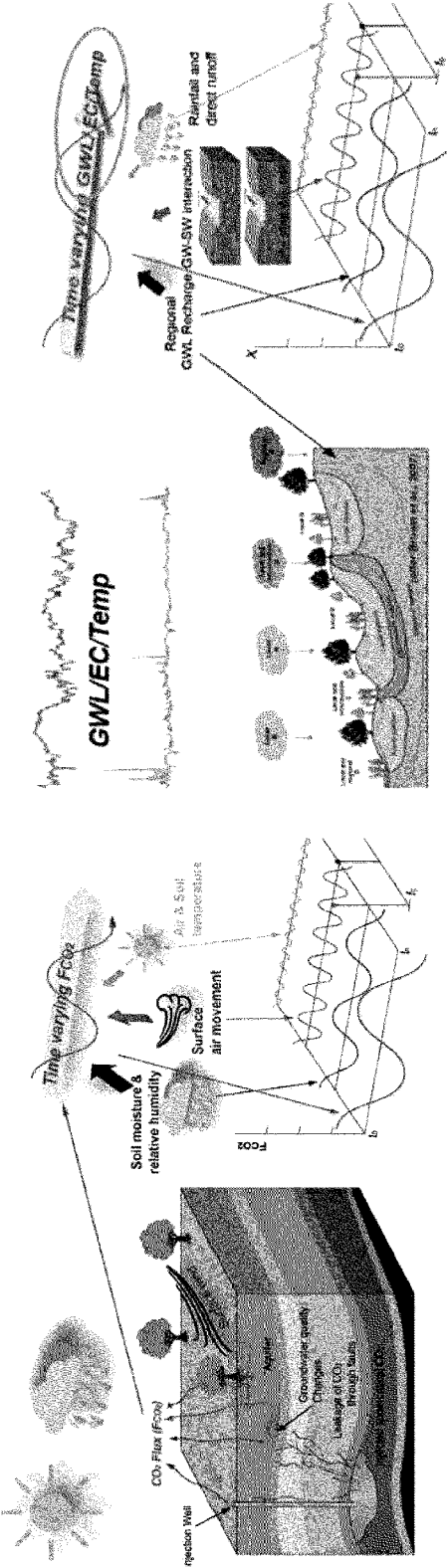
상기 구성된 기반데이터세트에 대한 시간영역 해상도에 따른 상기 다해상도 웨이블릿 해석을 통해 시간-주파수 영역에 대한 웨이블릿 에너지 분포 데이터를 도출하고, 상기 도출된 웨이블릿 에너지 분포 데이터에 대하여 상기 차원 축소 기법을 적용하여 잠재 환경 구동력 후보군을 선정하는 단계 및
상기 선정된 잠재 환경 구동력 후보군의 시간영역 해상도별 시간변화에 따른 변동특성을 추출하고, 상기 추출된 변동특성에 기초하여 상기 동적 특성을 추출 및 정량화하는 단계를 더 포함하는 훈련지표 최적화 방법.

[청구항 16] 제15항에 있어서,
상기 핵심 특성치 그룹을 선정하는 단계는,
상기 잠재 환경 구동력 후보군의 잠재 환경 구동력과, 상기 환경 계측 자료 간 다해상도 상관성을 판단하되, 상기 잠재 환경 구동력과 상기 관측자료간 시간지연 및 위상변화를 반영한 상관성 판단을 수행하고, 상기 수행된 상관성 판단 결과에 기초하여, 상기 잠재 환경 구동력과 상기 관측자료간 최대상관스케일을 선정하는 단계 및
상기 잠재 환경 구동력과 상기 관측자료간 웨이블릿 에너지 비율과 상기 선정된 최대상관스케일의 상관관계를 이용한 구동력을 식별하고, 상기 선정된 최대상관스케일의 결합에너지비율과 차원축소모델의 설명력 지표간의 선형결합을 처리하여 상대적 기여도를 평가하며, 상기 평가된 상대적 기여도에 기초하여 상기 핵심 특성치 그룹을 선정하는 단계를 더 포함하는 훈련지표 최적화 방법.

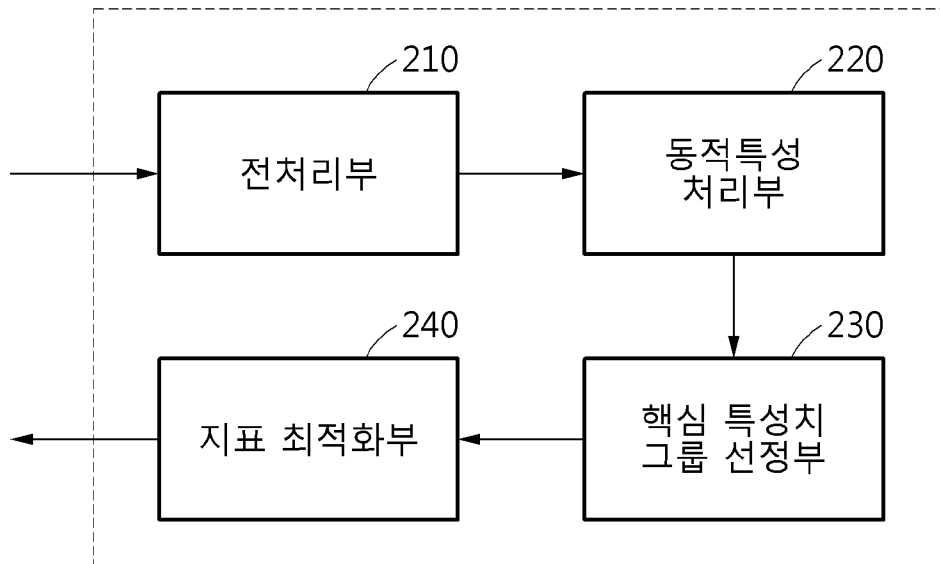
[청구항 17] 제14항에 있어서,
상기 복수의 훈련지표 값을 제어하는 단계는,
상기 핵심 특성치 그룹과 상기 환경 계측 자료를 입력으로 하는 장단기기억네트워크(Long-short Term Memory Network) 모델을 구축하고, 상기 핵심 특성치 그룹의 시간-주파수 영역에 기초하여 상기 복수의 훈련지표 값을 사전 정량화(pre-quantification)하는 단계;
상기 환경 계측 자료로부터 측정되는 관측값과, 상기 장단기기억네트워크 모델로부터 예측되는 예측값의 복합모델검증지표에 기초한 잔차검증 및 잔차의 다해상도해석에 기반하여 적어도 하나 이상의 예측 모델을 선정하고, 상기 선정된 적어도 하나 이상의 예측 모델 중 어느 하나의 예측 모델 또는 둘 이상의 예측 모델이 결합한 결합 예측 모델(combined prediction model)에 기초하여 상기 사전 정량화된 복수의 훈련지표를 사후 정량화(post-quantification)하는 단계 및
상기 정량화된 복수의 훈련지표 중 적어도 둘 이상의 훈련지표를 이용하는 최적훈련지표모형을 구성하는 단계

를 더 포함하는 훈련지표 최적화 방법.
[청구항 18] 제14항에 있어서,
상기 복수의 훈련지표 값은,
훈련기간(training period, T), 최소 배치 사이즈(minibatch size, mbs),
은닉층(hidden layers, HL)의 개수 및 최적 반복회수(epochs, E) 중 적어도
하나의 값을 포함하는
훈련지표 최적화 방법.

[도1]

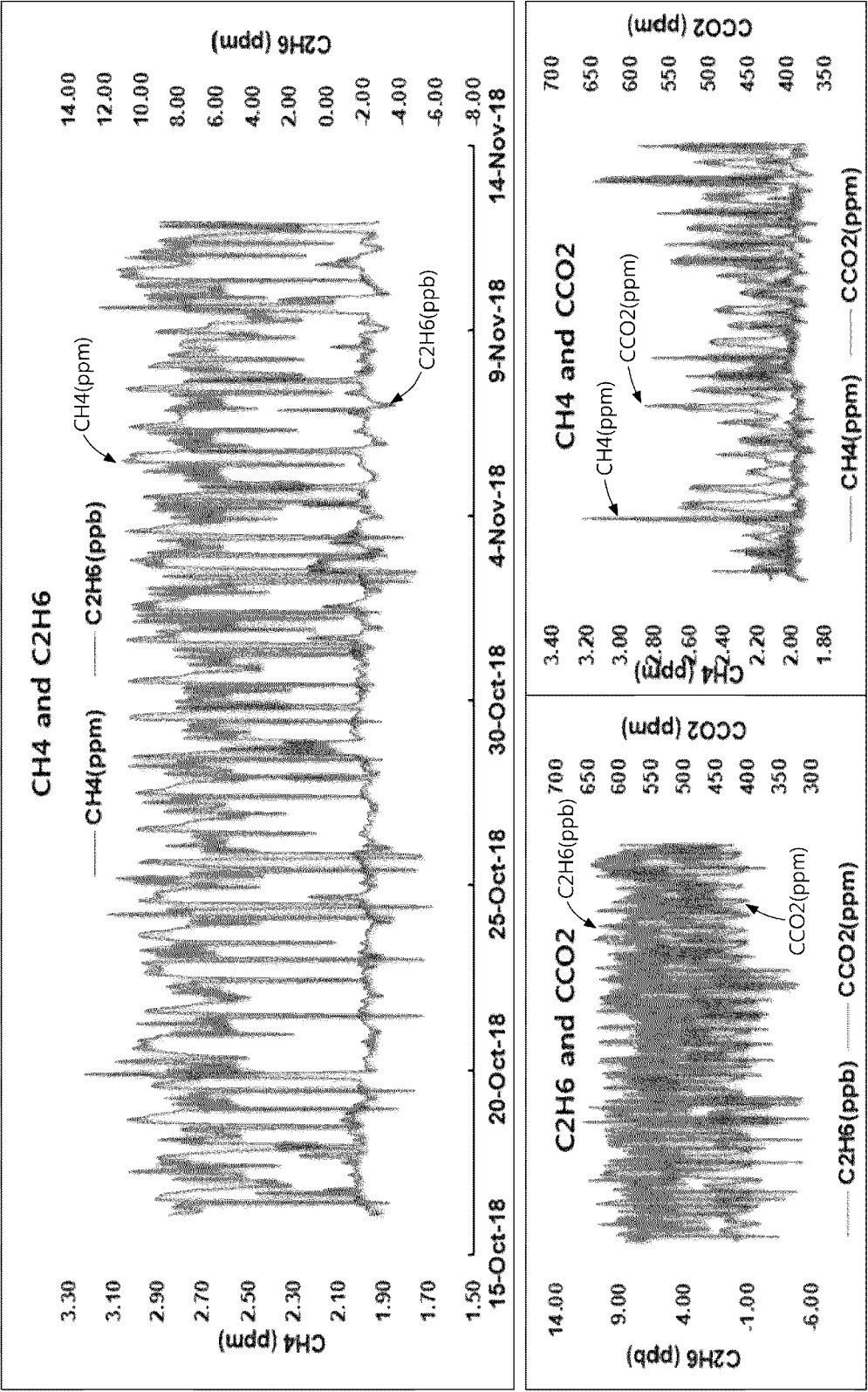


[도2]

200

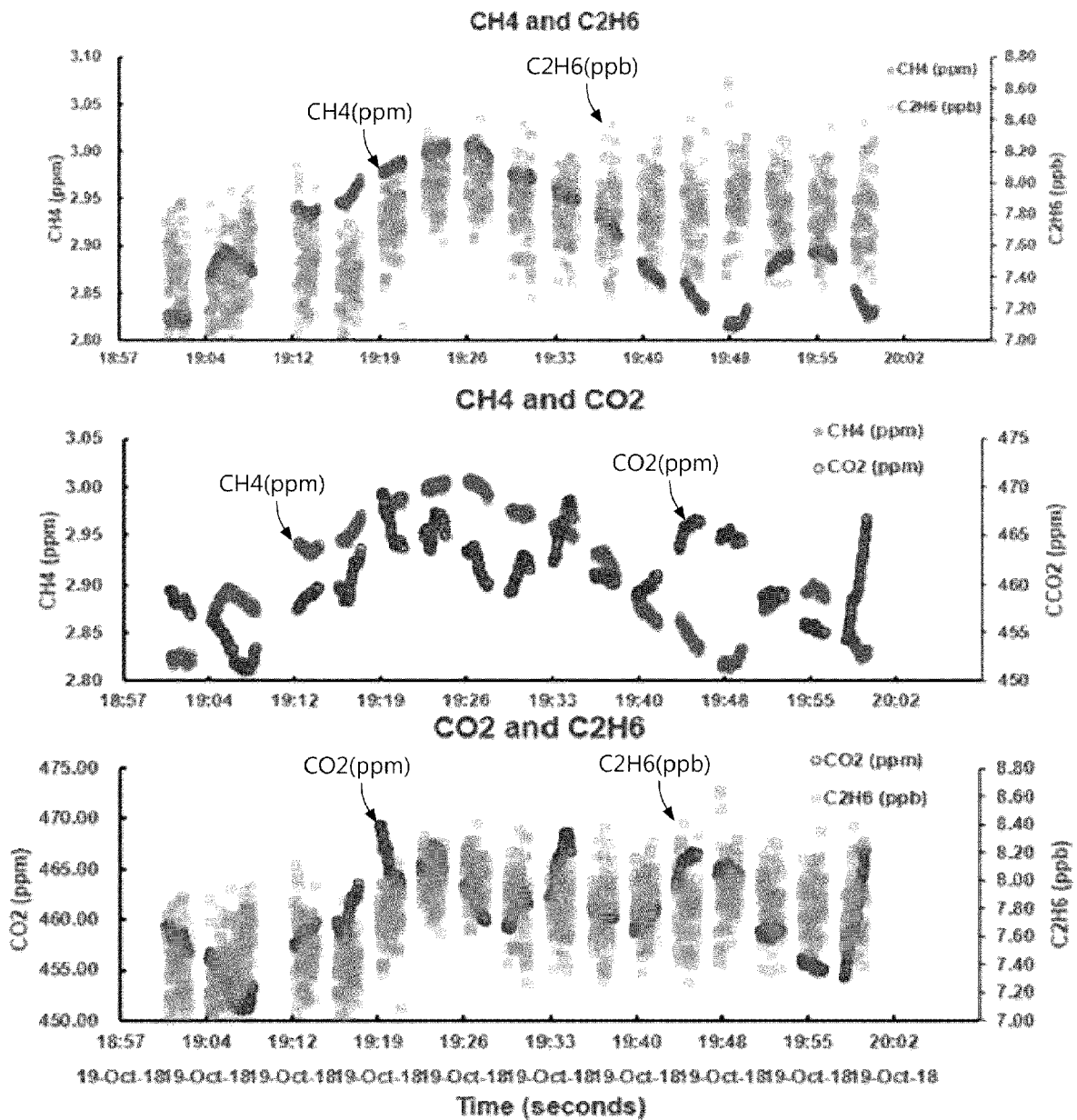
[도3a]

310



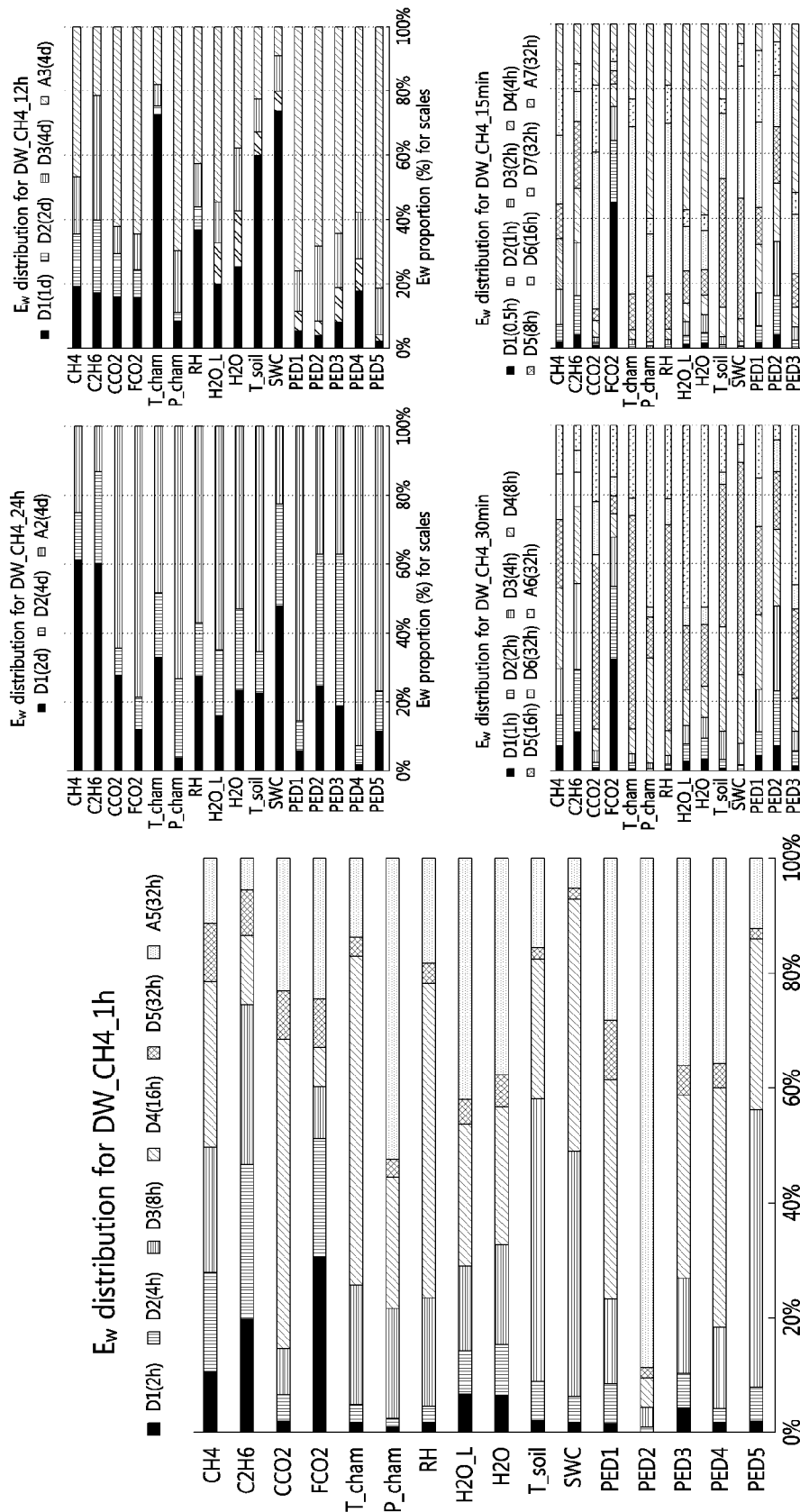
[도3b]

320



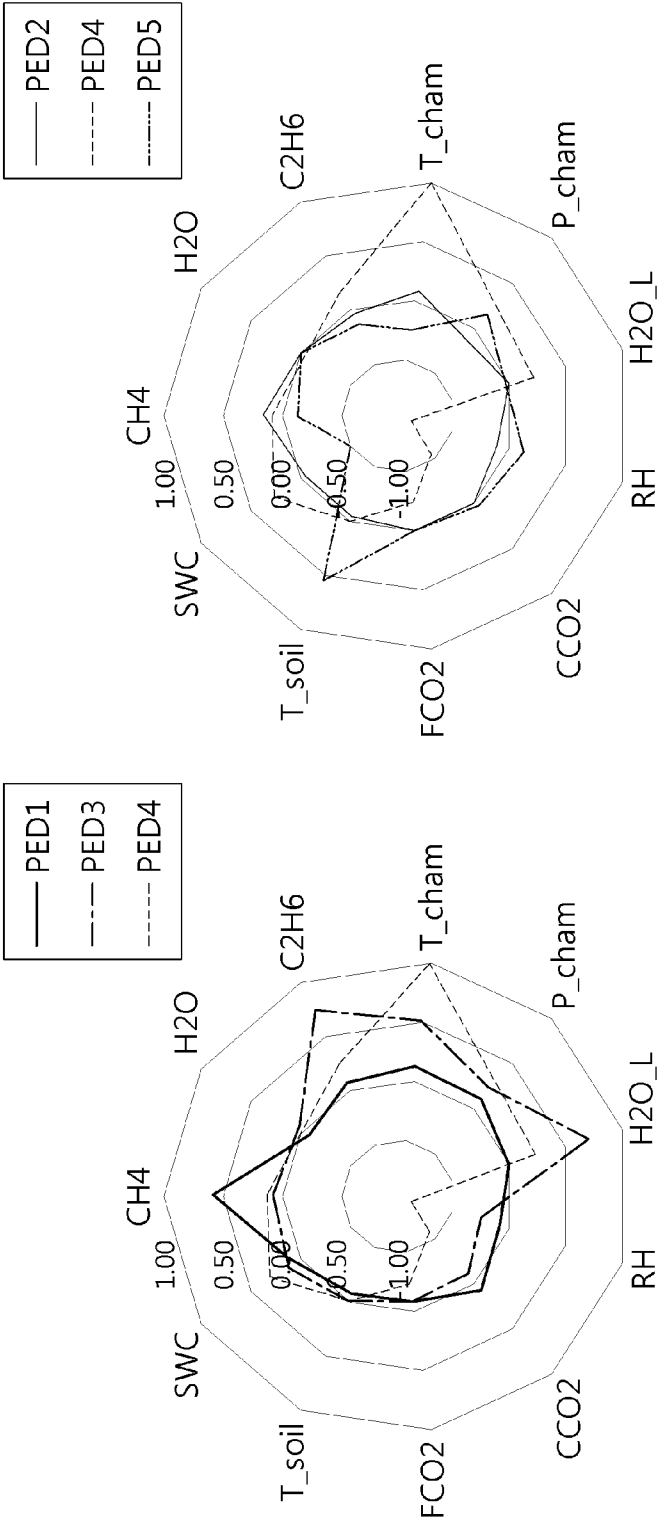
[도3c]

330



[도3d]

340

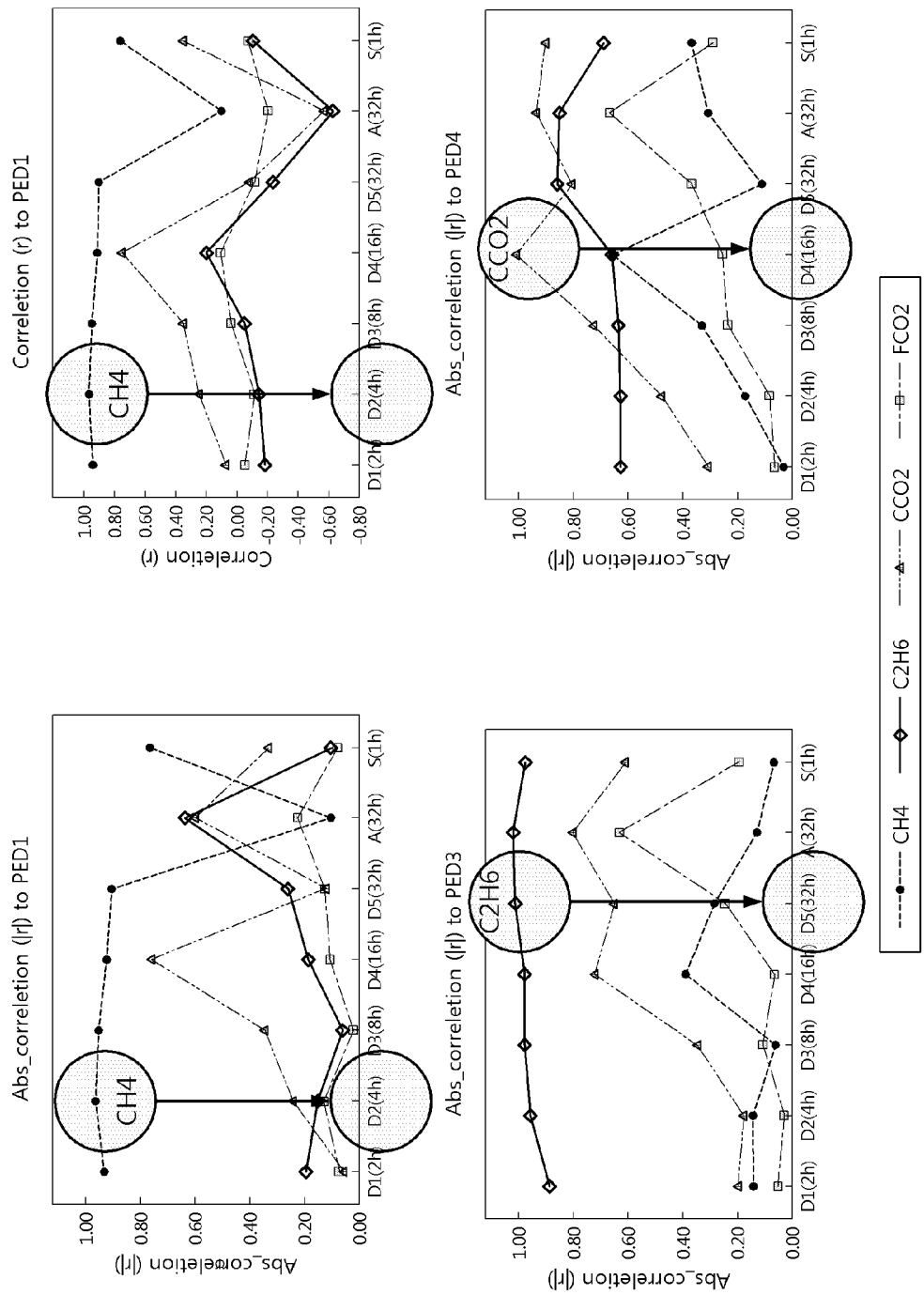


PEDs for soil gases

PEDs for environmental variables

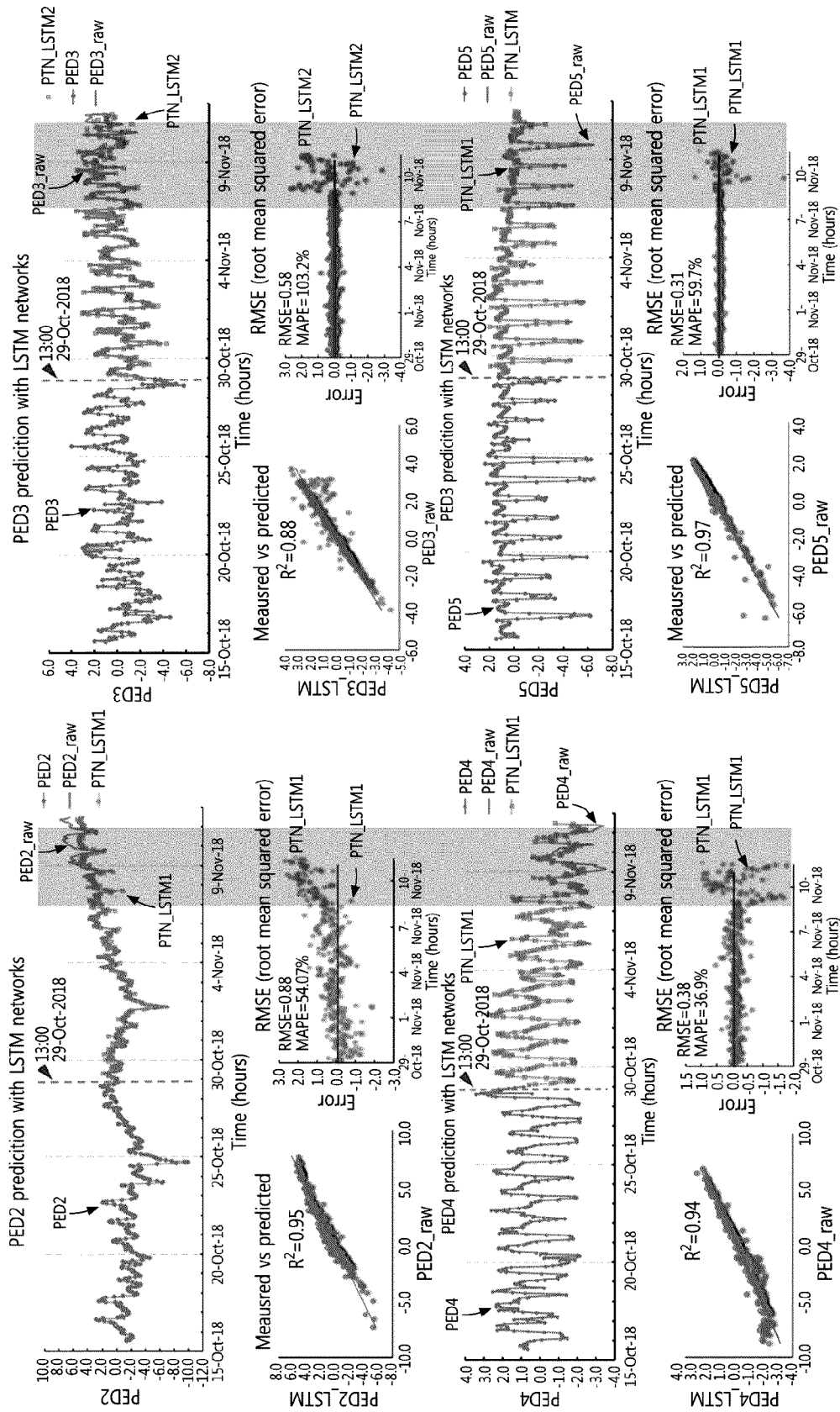
[도3e]

350



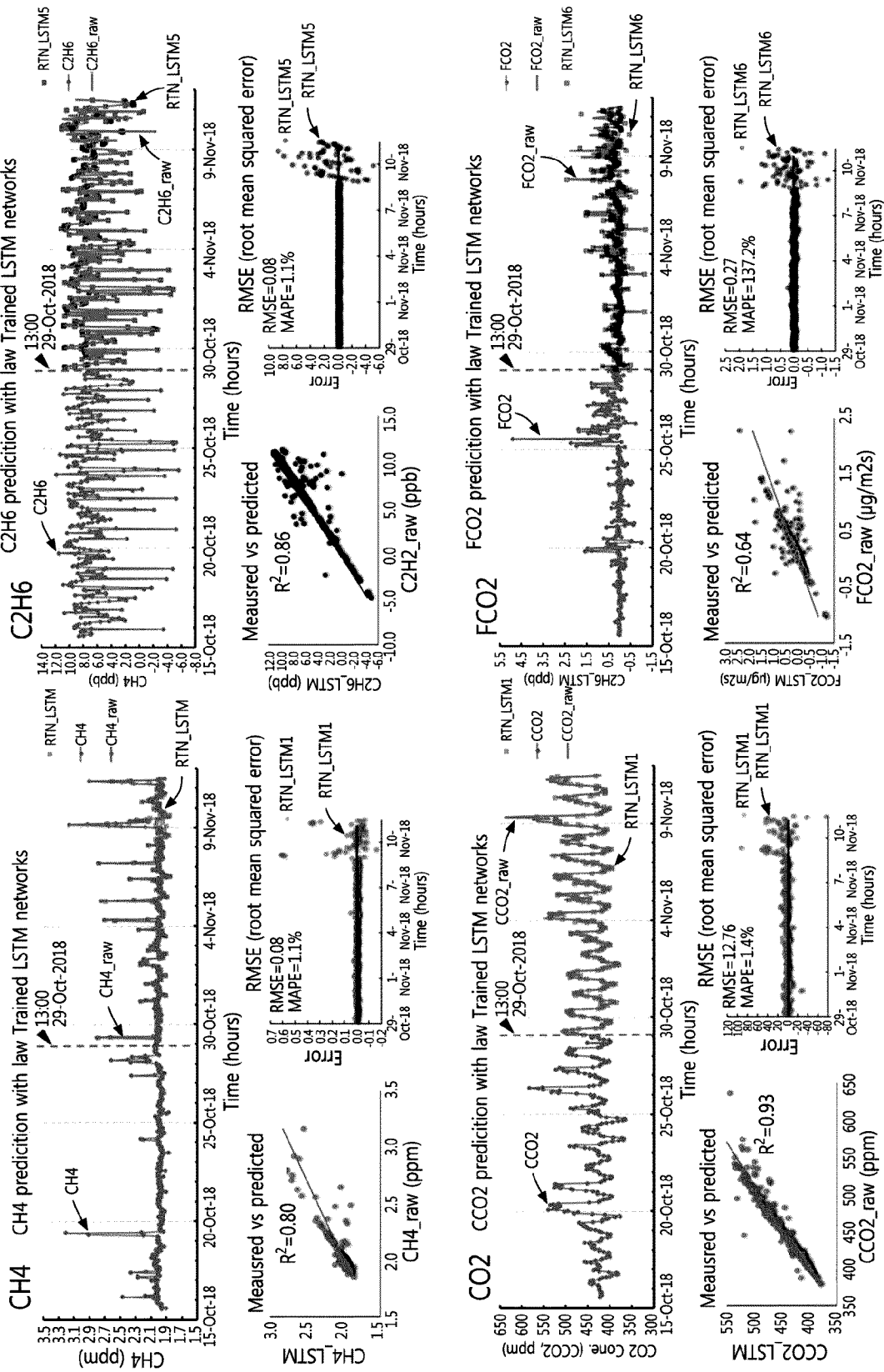
[도3f]

360



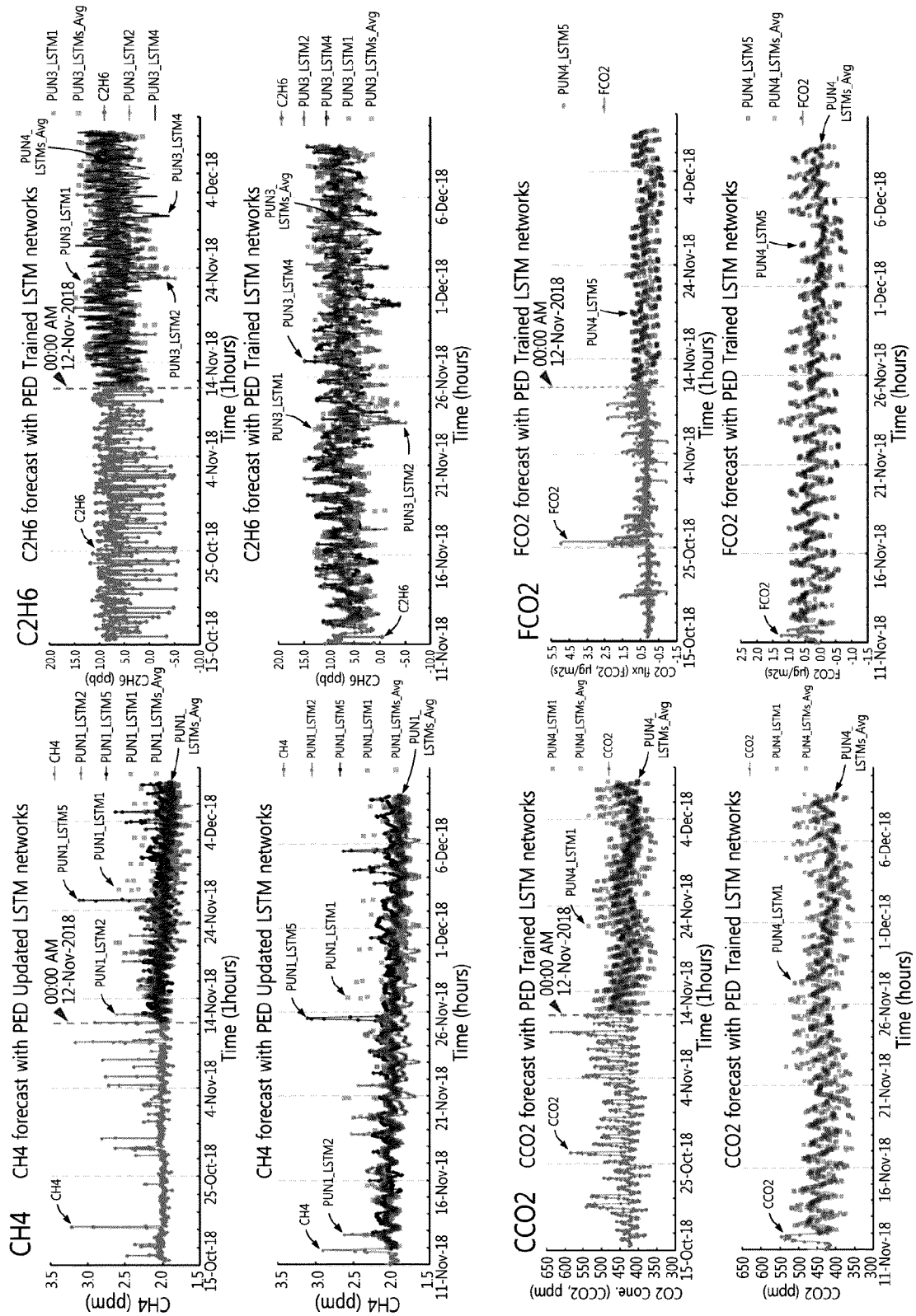
[도3g]

370



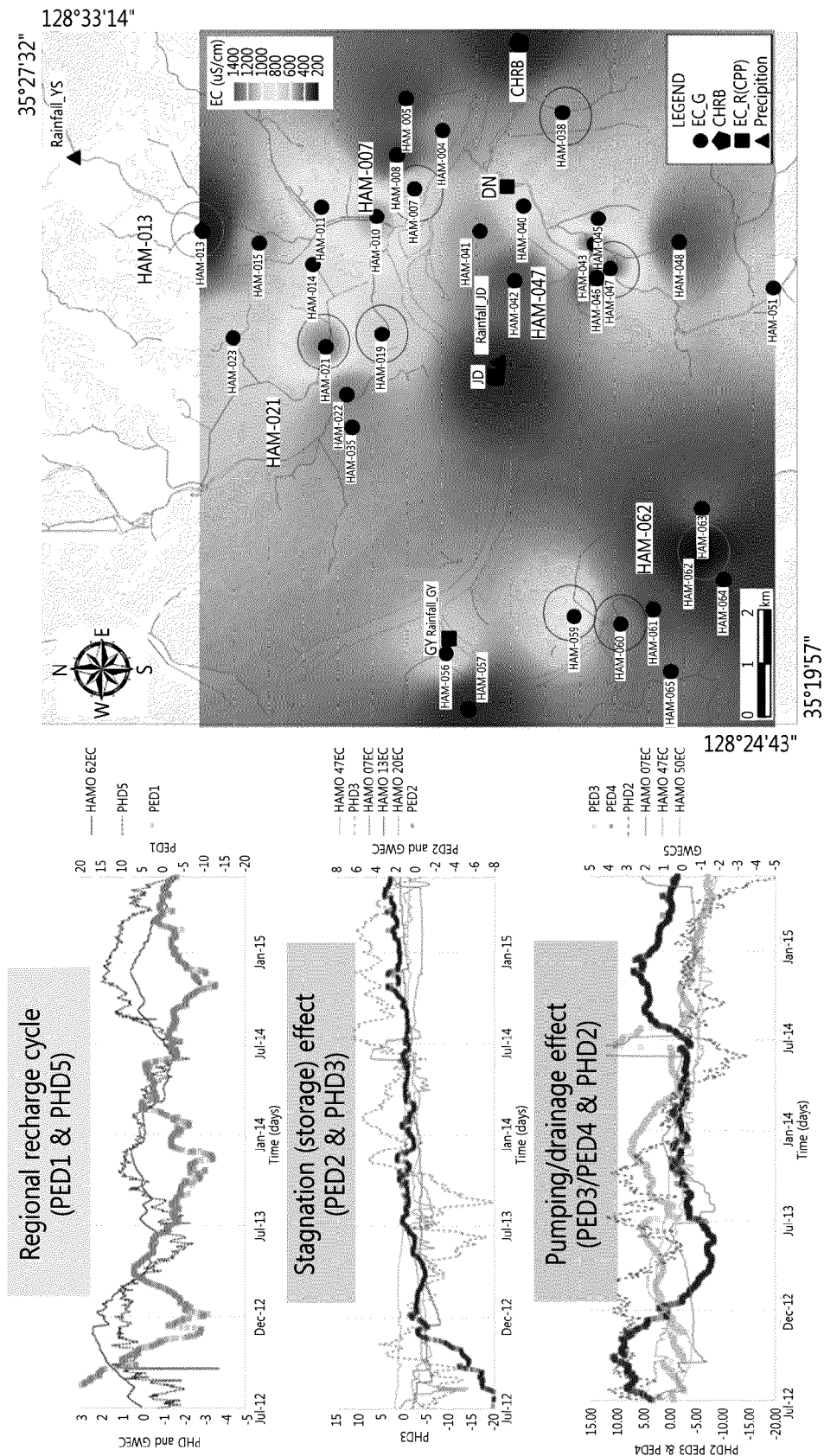
[도3h]

380



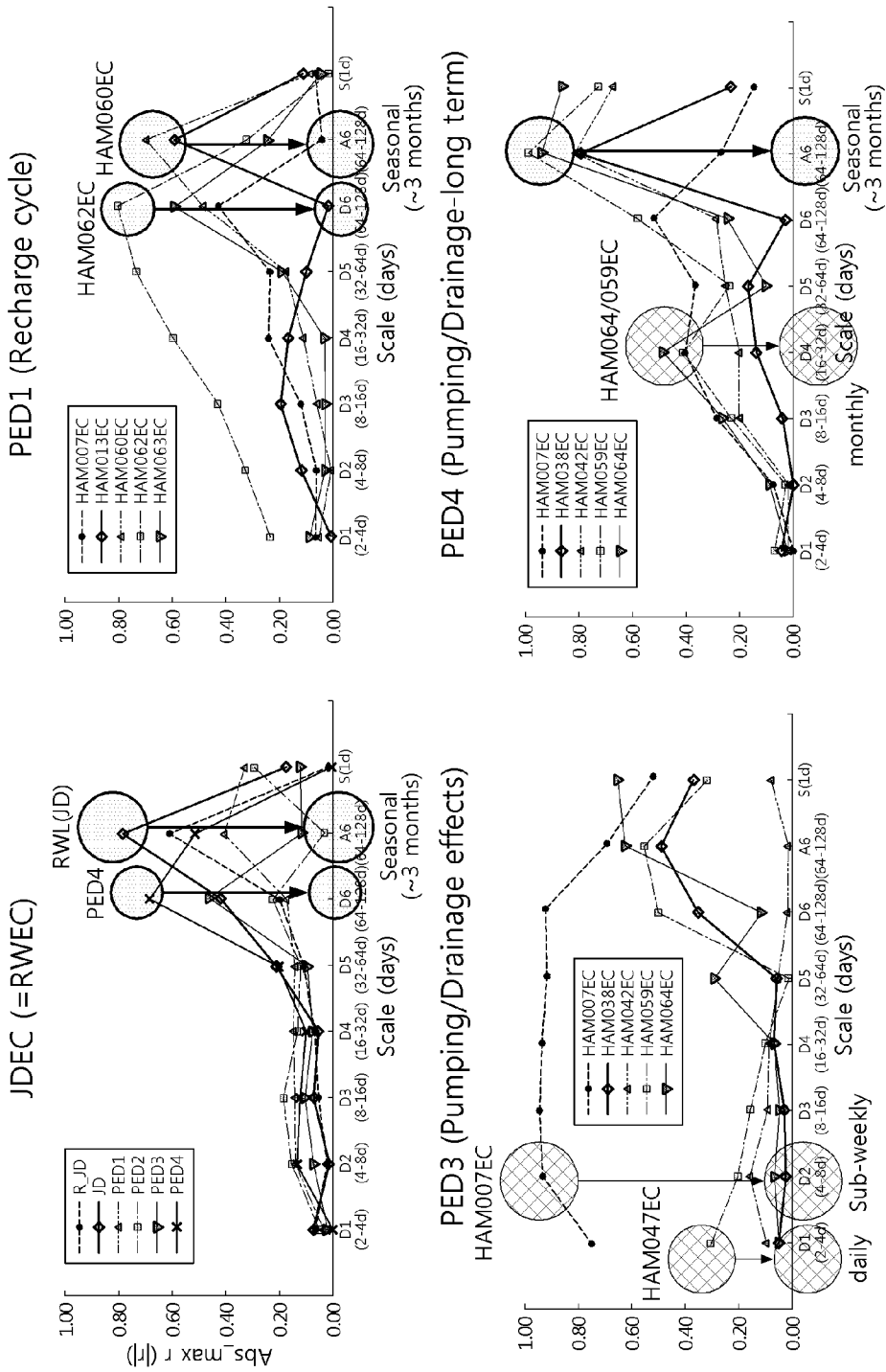
[도4b]

420

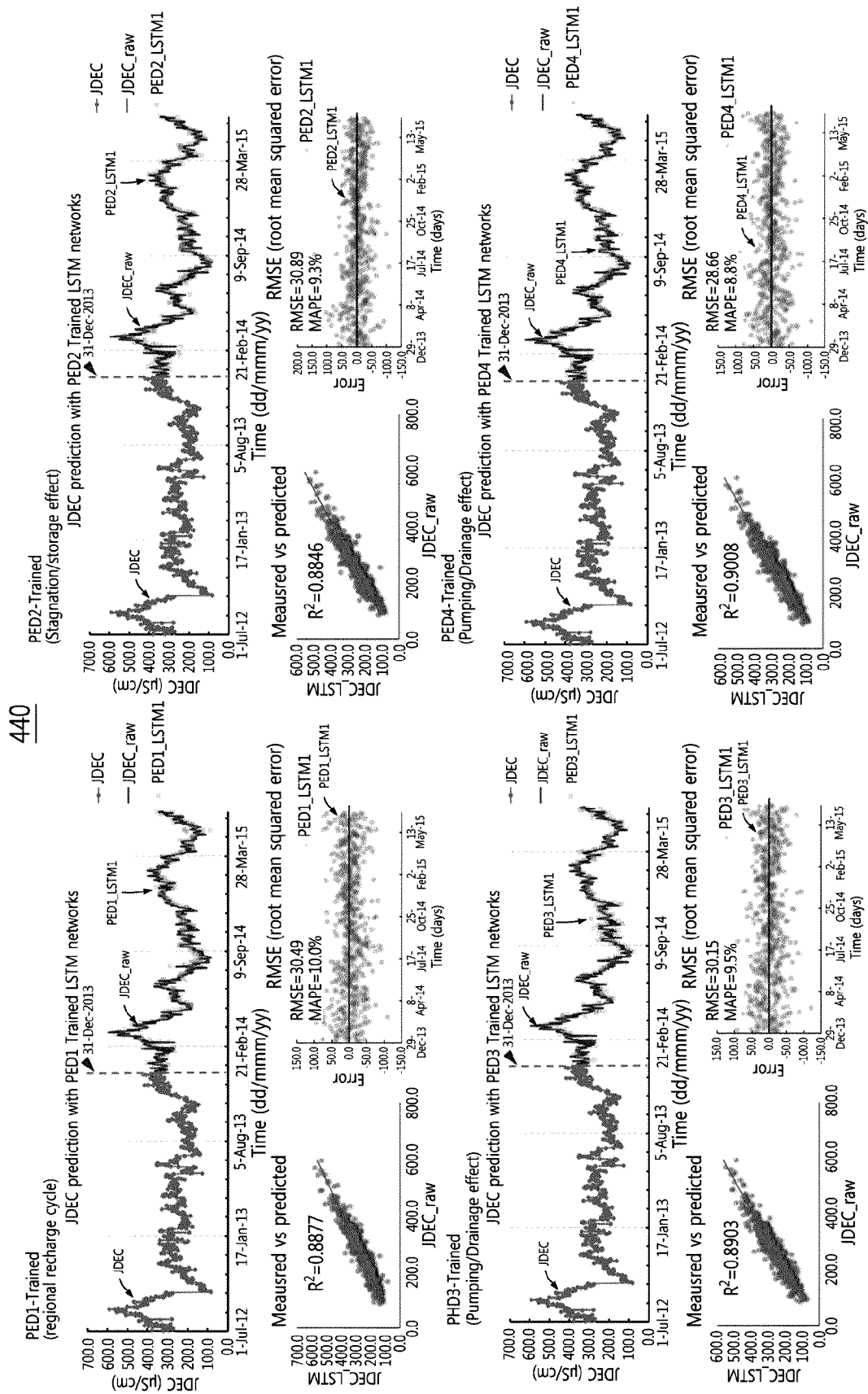


[도4c]

430

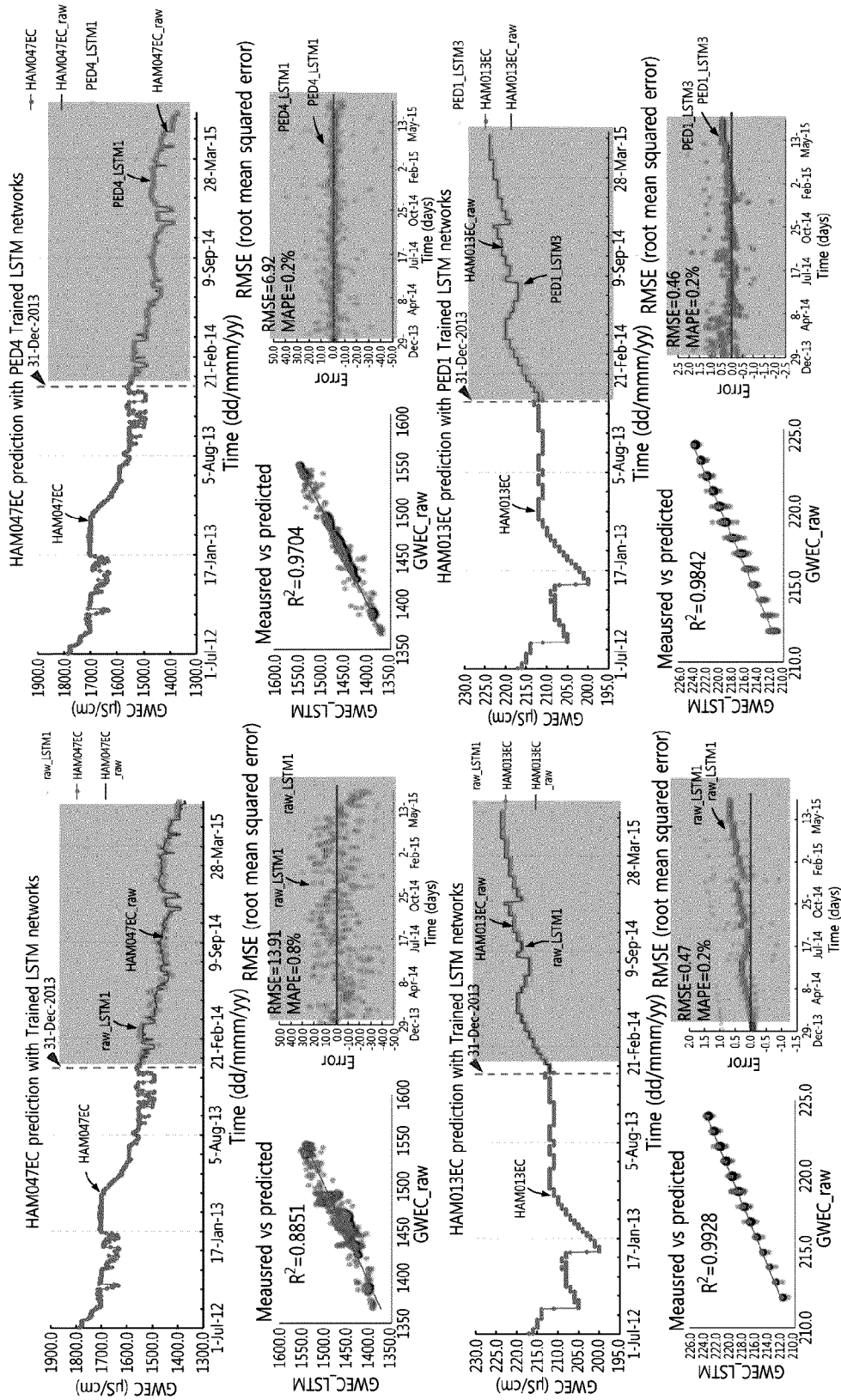


[도4d]



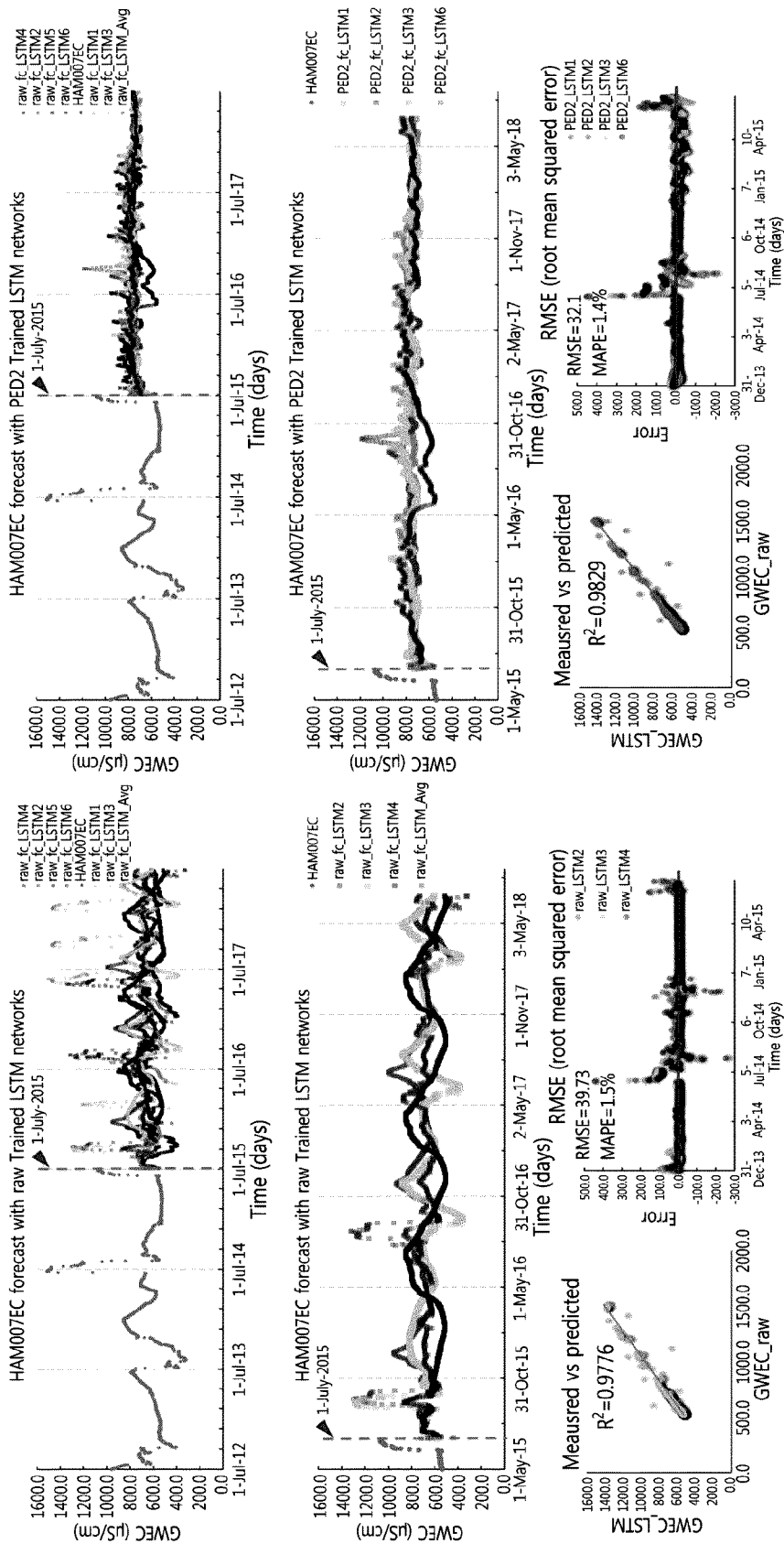
[도4e]

450



[도4f]

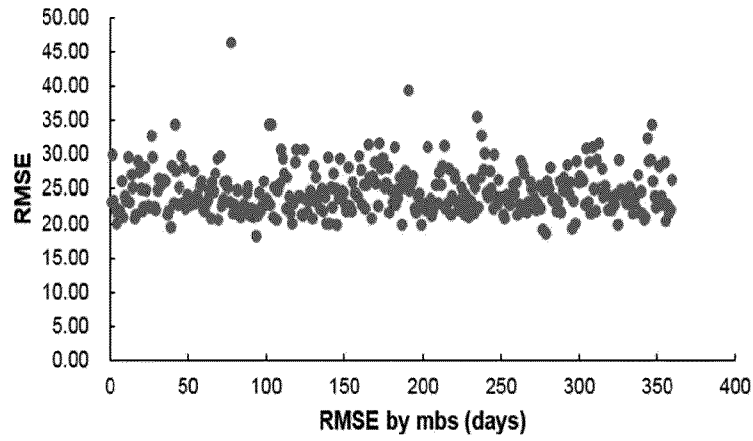
460



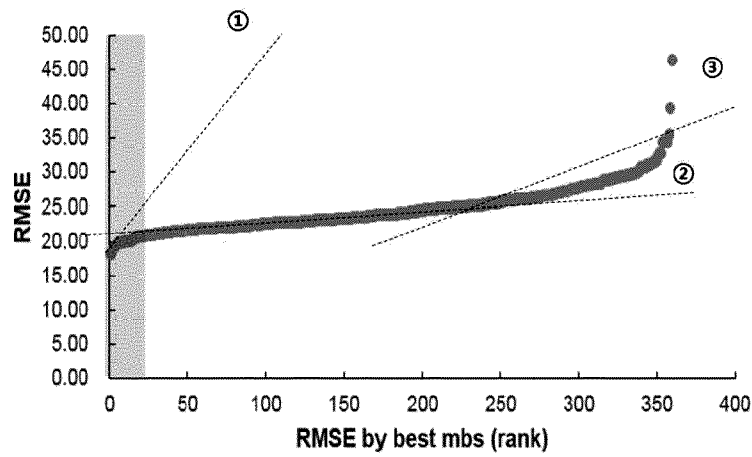
[도5a]

510

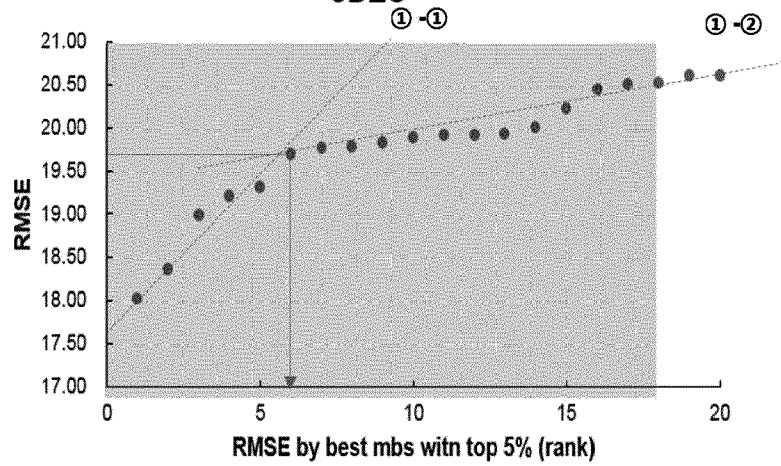
JDEC



JDEC

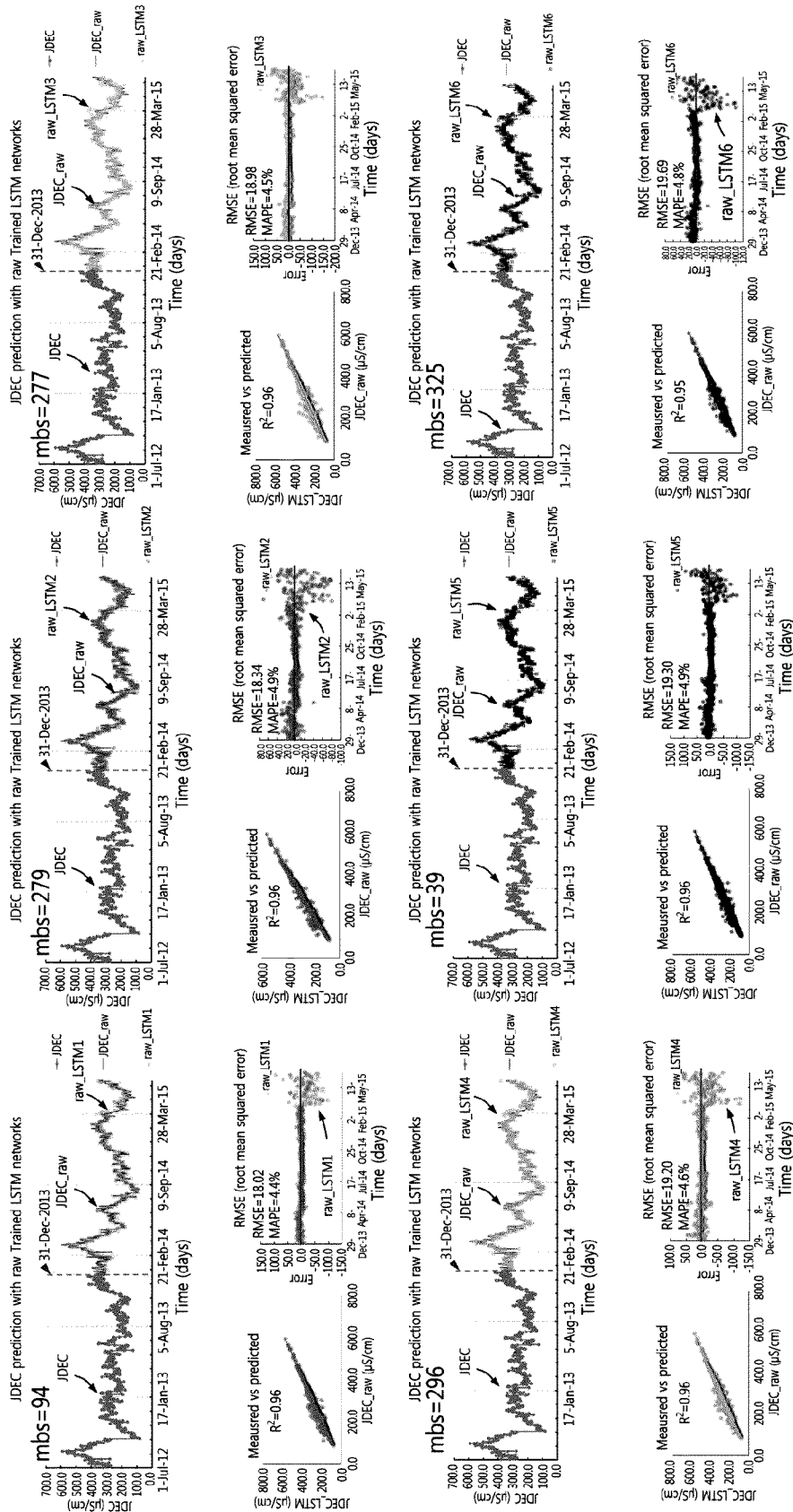


JDEC

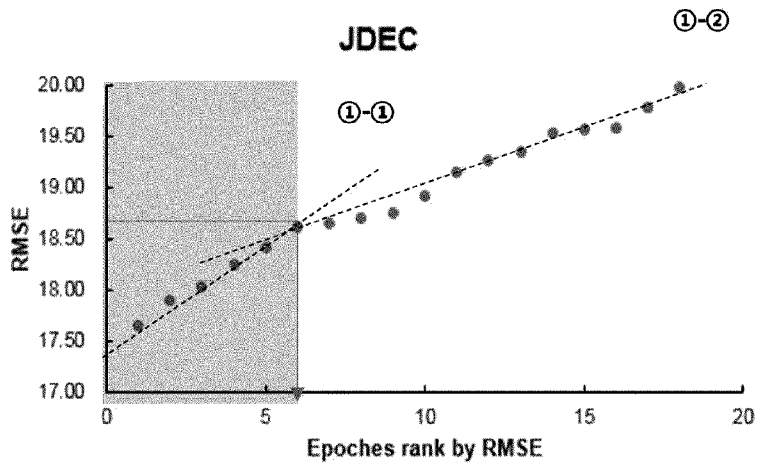
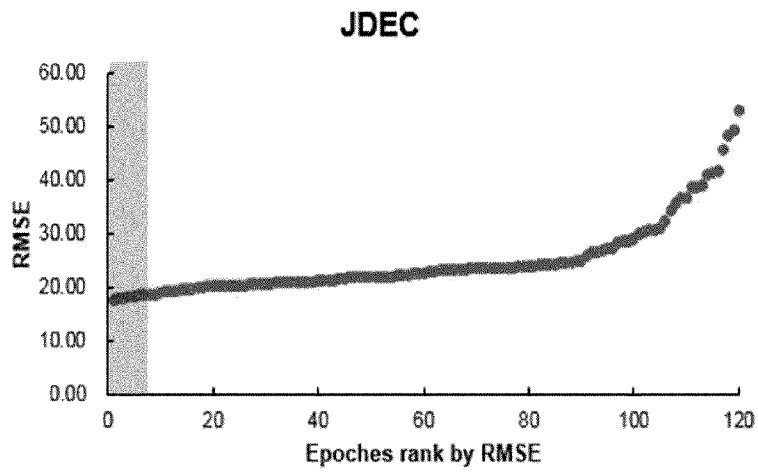
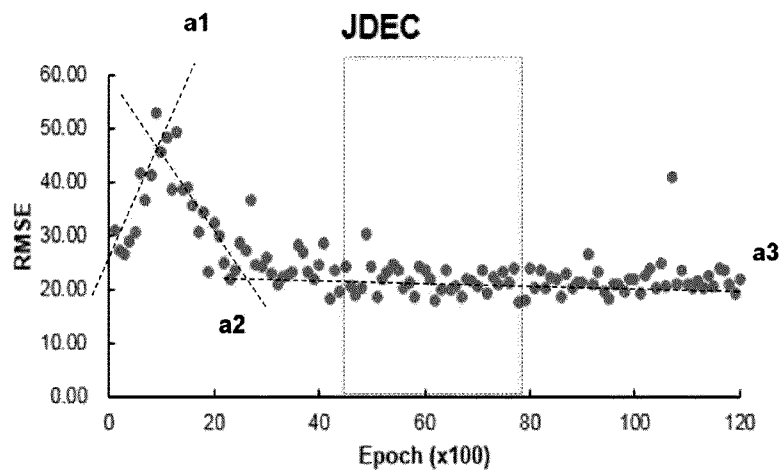


[도5b]

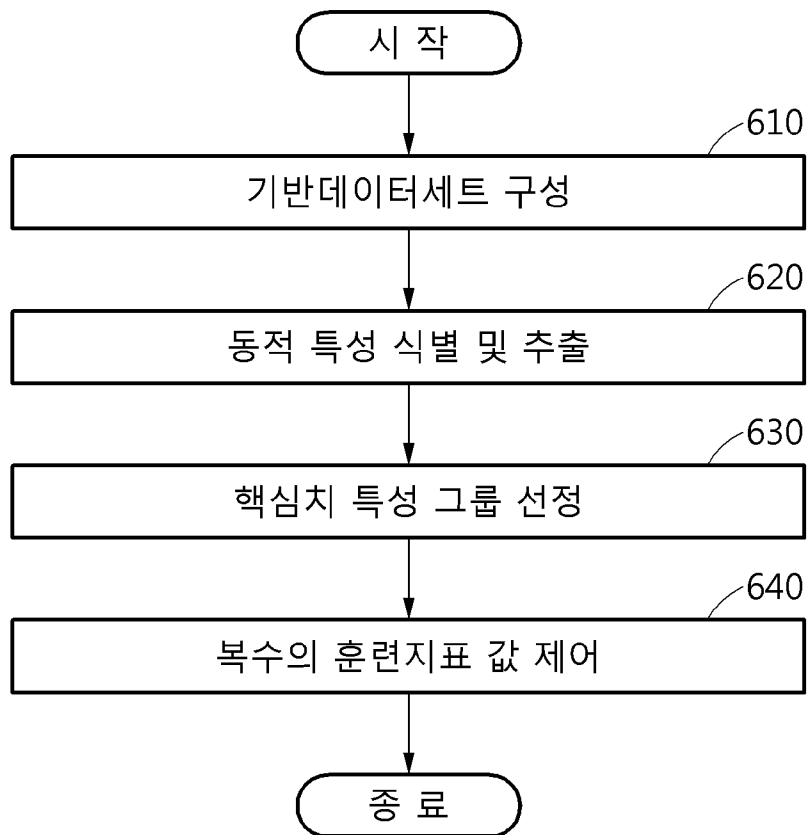
520



[도5c]

530

[도6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/011052

A. CLASSIFICATION OF SUBJECT MATTER G06N 3/08(2006.01)i; G06N 3/04(2006.01)i; G01W 1/10(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06N 3/08; G06N 3/02; G06N 3/04; G06N 99/00; G06Q 10/06; G01W 1/10 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models: IPC as above Japanese utility models and applications for utility models: IPC as above Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS (KIPO internal) & keywords: 핵심 특성치 그룹(key features group), 훈련지표(training indicator), 환경 예측 모델(environment prediction model), 최적화(optimization)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2018-0084969 A (FUJITSU LIMITED) 25 July 2018. See paragraphs [0052]-[0066]; and figures 2-4.	1-18
A	JP 2018-169959 A (JSOL CORP.) 01 November 2018. See paragraphs [0021]-[0030]; and figures 1-3.	1-18
A	KR 10-1959376 B1 (BAIDU USA LLC) 04 July 2019. See paragraphs [0016]-[0022]; and figure 1.	1-18
A	KR 10-2019-0050698 A (SAMSUNG ELECTRONICS CO., LTD.) 13 May 2019. See paragraphs [0019]-[0023]; and figure 1.	1-18
A	US 2017-0243140 A1 (DATAROBOT, INC.) 24 August 2017. See paragraphs [0066]-[0087]; and figure 1.	1-18
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 29 October 2020		Date of mailing of the international search report 30 October 2020
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex Daejeon Building 4, 189, Cheongsaro, Seo-gu, Daejeon, Republic of Korea 35208 Facsimile No. +82-42-481-8578		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2020/011052

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2018-0084969	A	25 July 2018	CN	108140144	A	08 June 2018
				JP	2019-508803	A	28 March 2019
				JP	6601569	B2	06 November 2019
				KR	10-2161902	B1	05 October 2020
				WO	2017-166155	A1	05 October 2017
JP	2018-169959	A	01 November 2018	JP	6458072	B2	23 January 2019
KR	10-1959376	B1	04 July 2019	CN	106875013	A	20 June 2017
				CN	106875013	B	02 June 2020
				EP	3179415	A1	14 June 2017
				EP	3179415	B1	18 September 2019
				JP	2017-107568	A	15 June 2017
				JP	6348561	B2	27 June 2018
				US	2017-0169326	A1	15 June 2017
KR	10-2019-0050698	A	13 May 2019	CN	109754078	A	14 May 2019
				US	2019-0138896	A1	09 May 2019
US	2017-0243140	A1	24 August 2017	AU	2017-345796	A1	23 May 2019
				EP	3529755	A1	28 August 2019
				GB	2541625	A	22 February 2017
				GB	2571651	A	04 September 2019
				JP	2017-520068	A	20 July 2017
				JP	2019-023906	A	14 February 2019
				JP	2019-023907	A	14 February 2019
				JP	2019-537125	A	19 December 2019
				JP	6444494	B2	26 December 2018
				JP	6457693	B1	23 January 2019
				KR	10-2019-0108559	A	24 September 2019
				US	10366346	B2	30 July 2019
				US	10496927	B2	03 December 2019
				US	10558924	B2	11 February 2020
				US	2015-0339572	A1	26 November 2015
				US	2016-0335550	A1	17 November 2016
				US	2016-0364647	A1	15 December 2016
				US	2018-0046926	A1	15 February 2018
				US	2018-0060738	A1	01 March 2018
				US	2018-0060744	A1	01 March 2018
				US	2020-0090075	A1	19 March 2020
				US	2020-0134489	A1	30 April 2020
				US	9489630	B2	08 November 2016
				US	9652714	B2	16 May 2017
				US	9659254	B2	23 May 2017
				WO	2015-179778	A1	26 November 2015
				WO	2018-075995	A1	26 April 2018

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06N 3/08(2006.01)i, G06N 3/04(2006.01)i, G01W 1/10(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06N 3/08; G06N 3/02; G06N 3/04; G06N 99/00; G06Q 10/06; G01W 1/10

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 핵심 특성치 그룹(key features group), 훈련지표(training indicator), 환경 예측 모델(environment prediction model), 최적화(optimization)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-2018-0084969 A (후지쯔 가부시끼가이샤) 2018.07.25 단락 [0052]-[0066]; 및 도면 2-4	1-18
A	JP 2018-169959 A (JSOL CORP.) 2018.11.01 단락 [0021]-[0030]; 및 도면 1-3	1-18
A	KR 10-1959376 B1 (바이두 유에스에이 엘엘씨) 2019.07.04 단락 [0016]-[0022]; 및 도면 1	1-18
A	KR 10-2019-0050698 A (삼성전자주식회사) 2019.05.13 단락 [0019]-[0023]; 및 도면 1	1-18
A	US 2017-0243140 A1 (DATAROBOT, INC.) 2017.08.24 단락 [0066]-[0087]; 및 도면 1	1-18

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“D” 본 국제출원에서 출원인이 인용한 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후 “X”에 공개된 선출원 또는 특허 문헌

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2020년 10월 29일 (29.10.2020)

국제조사보고서 발송일

2020년 10월 30일 (30.10.2020)

ISA/KR의 명칭 및 우편주소



대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

양정록

전화번호 +82-42-481-5709



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2018-0084969 A	2018/07/25	CN 108140144 A JP 2019-508803 A JP 6601569 B2 KR 10-2161902 B1 WO 2017-166155 A1	2018/06/08 2019/03/28 2019/11/06 2020/10/05 2017/10/05
JP 2018-169959 A	2018/11/01	JP 6458072 B2	2019/01/23
KR 10-1959376 B1	2019/07/04	CN 106875013 A CN 106875013 B EP 3179415 A1 EP 3179415 B1 JP 2017-107568 A JP 6348561 B2 US 2017-0169326 A1	2017/06/20 2020/06/02 2017/06/14 2019/09/18 2017/06/15 2018/06/27 2017/06/15
KR 10-2019-0050698 A	2019/05/13	CN 109754078 A US 2019-0138896 A1	2019/05/14 2019/05/09
US 2017-0243140 A1	2017/08/24	AU 2017-345796 A1 EP 3529755 A1 GB 2541625 A GB 2571651 A JP 2017-520068 A JP 2019-023906 A JP 2019-023907 A JP 2019-537125 A JP 6444494 B2 JP 6457693 B1 KR 10-2019-0108559 A US 10366346 B2 US 10496927 B2 US 10558924 B2 US 2015-0339572 A1 US 2016-0335550 A1 US 2016-0364647 A1 US 2018-0046926 A1 US 2018-0060738 A1 US 2018-0060744 A1 US 2020-0090075 A1 US 2020-0134489 A1 US 9489630 B2 US 9652714 B2 US 9659254 B2 WO 2015-179778 A1 WO 2018-075995 A1	2019/05/23 2019/08/28 2017/02/22 2019/09/04 2017/07/20 2019/02/14 2019/02/14 2019/12/19 2018/12/26 2019/01/23 2019/09/24 2019/07/30 2019/12/03 2020/02/11 2015/11/26 2016/11/17 2016/12/15 2018/02/15 2018/03/01 2018/03/01 2020/03/19 2020/04/30 2016/11/08 2017/05/16 2017/05/23 2015/11/26 2018/04/26