



CHALMERS



Modellering av Sjöbrisen i Los Angeles

Kandidatarbete MMSX21-VT26-46A

Kandidatarbete inom Program Maskinteknik

OLLE ARONSSON
MOA ÖHMAN
EBBA BERNTSSON

INSTITUTIONEN FÖR MEKANIK OCH MARITIMA VETENSKAPER

CHALMERS TEKNISKA HÖGSKOLA

Göteborg, Sverige 2026

www.chalmers.se

KANDIDATARBETE I MASKINTEKNIK OCH KEMITEKNIK MED FYSIK

Modellering av sjöbrisen i Los Angeles

Kandidatarbete MMSX21-VT26-46A

OLLE ARONSSON
MOA ÖHMAN
EBBA BERNTSSON



CHALMERS

Institutionen för Mekanik och Maritima Vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
Göteborg 2026

Modellering av sjöbrisen i Los Angeles
Kandidatarbete MMSX21-VT26-46A
OLLE ARONSSON
MOA ÖHMAN
EBBA BERNTSSON

© , OLLE ARONSSON, MOA ÖHMAN,
EBBA BERNTSSON 2026.

Handledare: Thisal Mandula Sugathapala, Institutionen för Mekanik och Maritima Vetenskaper
Examinator: Niklas Andersson, Institutionen för Mekanik och Maritima Vetenskaper

Kandidatarbete 2026
Institutionen för Mekanik och Maritima Vetenskaper
Chalmers Tekniska Högskola
SE-412 96 Göteborg
Telefon +46 31 772 1000

Omslagsbild: Bild tagen under ett race i seglingsklassen 49er under världscupen i Hyeres 2026.

Typsatt i L^AT_EX
Göteborg 2026

Sammandrag

Sjöbris är en mesoskalig atmosfärisk cirkulation som uppstår till följd av temperaturgradienter mellan land och hav och har en betydande inverkan på lokala vindförhållanden i kustnära områden. Trots betydelsen av tillförlitliga vindprognoser inför seglingstävlingarna under Olympiska sommarspelen i Los Angeles 2028 är kunskapen om lokala sjöbrismönster och prognosmodeller för området utanför San Pedro fortfarande begränsad. Syftet med studien är att öka förståelsen för de fysikaliska mekanismer som styr sjöbriscirkulationen samt att utveckla och utvärdera en modell för att identifiera och prognostisera sjöbris i området utanför San Pedro, Los Angeles.

Den framtagna modellen tränas på reanalysdata från ERA5, ett globalt dataset som kombinerar observationer och numerisk modellering för att kunna återge atmosfärens tillstånd. Under prognostiseringen för nästkommande dag används prognosdata från Global Forecast System (GFS), en global numerisk prognosmodell. En sekvensbaserad maskininlärningsmodell av typen Long Short-Term Memory (LSTM) används för att identifiera temporala beroenden i atmosfäriska processer baserat på relevanta fysiska variabler.

Resultaten visar att modellen kan identifiera förekomsten av sjöbris under en given dag samt uppskatta timvisa vindförhållanden i termer av vindriktning och vindhastighet. Modellens prestanda påverkas dock av skillnader i upplösning mellan träningsdata och observationsdata, vilket introducerar osäkerheter i prognosen. Studien indikerar att datadrivna modeller kan utgöra ett värdefullt komplement till traditionella prognosmetoder, men att ett noggrant val av data är avgörande för modellens tillförlitlighet.

Abstract

Sea breeze is a mesoscale atmospheric circulation that develops as a result of temperature gradients between land and sea and has a significant impact on local wind conditions in coastal areas. However, predictive data and localized forecasting models for sea breeze patterns in the San Pedro region ahead of the 2028 Olympic sailing competitions remain limited, despite the importance of accurate wind forecasts for race preparation and tactical decision-making. The aim of this study is to improve the understanding of the physical mechanisms influencing the sea breeze circulation, as well as to develop and evaluate a model for identifying and forecasting sea breeze events in the area off San Pedro, Los Angeles.

The developed model is trained using ERA5 reanalysis data, a global dataset that combines observations and numerical modelling to provide a consistent representation of the state of the atmosphere. For forecasting the following day, forecast data from the Global Forecast System (GFS), a global numerical weather prediction model, are used. A sequence-based machine learning model of the Long Short-Term Memory (LSTM) type is employed to identify temporal dependencies in atmospheric processes based on relevant physical variables.

The results show that the model is capable of identifying the occurrence of sea breeze during a given day and estimating hourly wind conditions in terms of wind direction and wind speed. However, the model's performance is affected by differences in resolution between the training data and the observational data, introducing uncertainties in the forecast. The study indicates that data-driven models can serve as a valuable complement to traditional forecasting methods, although careful selection of data is crucial for model reliability.

Innehåll

1	Inledning	1
1.1	Bakgrund	1
1.2	Syfte och mål	1
1.3	Uppgift	1
1.4	Avgränsningar	2
2	Teori	2
2.1	Sjöbris - översikt	3
2.1.1	Sjöbrisens uppbyggnad	3
2.1.2	Bjerknes cirkulations teori	4
2.2	Tryckfält	5
2.3	Bakgrundsvind	5
2.3.1	Sjöbrisindex	6
2.3.2	Olika typer av sjöbris	6
2.4	Jordens rotation	7
2.4.1	Corioliskraften	7
2.4.2	Corioliskraftens påverkan på vinden	8
2.5	Vertikala lufttransporter i gränsskiktet	9
2.6	Turbulens i atmosfäriska gränsskiktet	9
2.6.1	Reynoldsdekomposition av vindhastigheten	9
2.6.2	Turbulent transport av rörelsemängd	10
2.6.3	Turbulent transport av värme	10
2.6.4	Turbulent kinetisk energi	10
2.6.5	Turbulens i sjöbriscirkulation	11
2.7	Modeller	11
2.7.1	Analytiska modeller	11
2.7.2	Numeriska modeller	12
2.8	Maskininlärning	13
2.8.1	Deep Learning	14
2.8.2	PyTorch	15
2.9	Dataset	15
2.9.1	ERA5	15
2.9.2	GFS	16
2.10	Observationssystem och väderbojar	18
2.10.1	Väderbojar utanför San Pedro	18
3	Metod	19
3.1	Databearbetning	19
3.2	Konstruktion av egenskaper och målvariabler	21
3.2.1	Vindrepresentationer	21
3.2.2	Land-Hav gradienter	22
3.2.3	Tidsrepresentation	22
3.2.4	Dynamiska egenskaper	22
3.2.5	Sammansatta index	22
3.3	Regelbaserad klassificering av sjöbris	23
3.4	Prognoshorisonter och målvariabler	23
3.5	LSTM-modell	24

3.5.1	Sekvensbaserad indatarepresentation	24
3.5.2	Nätverksarkitektur	24
3.6	Modellens utdata	25
3.6.1	Binär klassificering av sjöbris	25
3.6.2	Regression av framtida vindförändringar	26
3.7	Förlustfunktion	26
3.7.1	Klassificeringsförlust	26
3.7.2	Regressionsförlust	27
3.7.3	Total förlust	27
3.8	Träning av modellen	27
3.9	Kalibrering och beslutströskel	28
3.9.1	Val av beslutströskel	28
3.10	Utvärderingsmetod	28
3.10.1	Precision	28
3.10.2	Recall	28
3.10.3	F1-score	29
3.10.4	Balanserad träffsäkerhet (Balanced accuracy)	29
3.10.5	Confusion matrix	29
3.11	Modellval under träning	29
3.12	Validering av regressionsutgångar	30
3.12.1	Vindriktning	30
3.13	Jämförelse mot persistensbaslinje	30
3.14	Fallstudie och operativ verifiering	31
3.15	Sammanfattning av valideringsmetod	31
4	Resultat	33
4.1	Träning av modellen	33
4.1.1	Klassificeringsresultat	34
4.1.2	Felmått för vindprognoser	35
4.1.3	Kalibrering av sannolikheter	35
4.2	Validering och verifiering	35
4.2.1	ROC- och precision-recall-analys	36
4.2.2	Validering mot NDBC-data	37
4.3	Fallstudie	38
4.3.1	Klassificeringsresultat för fallstudien	38
4.3.2	Vindprognos i fallstudien	39
4.3.3	Tidsserieanalys av fallstudien	39
4.4	Prognosvy för nästkommande dag	42
4.4.1	Sjöbrissignal	43
4.4.2	Vindhastighet	43
4.4.3	Vindriktning	44
4.5	Sammanfattning av resultat	44
5	Diskussion och analys	45
5.1	Modellens prestanda	45
5.1.1	Motivering av modellval	45
5.1.2	Utvärdering av träningen på modellen	45
5.1.3	Analys av Fallstudie	46
5.2	Fysikalisk tolkning av resultaten	47

5.2.1	Geografiska områden	47
5.2.2	Övriga kopplingar till teorin	48
5.3	Begränsningar och Osäkerhet	49
5.3.1	Dataupplösning	49
5.3.2	Modellens begränsning	49
5.3.3	Klimatförändringens inverkan på framtidens prognoser	50
5.4	Operativ användbarhet	50
6	Slutsats	52
	Referenser	53

Nomenklatur

Symboler

Nedan presenteras de symboler och förkortningar som förekommer återkommande i rapporten.

Symbol	Enhet	Beskrivning
$\bar{U}$	m s^{-1}	Medelvindhastighet i öst-västlig riktning
$\bar{V}$	m s^{-1}	Medelvindhastighet i nord-sydlig riktning
p	Pa	Lufttryck
ρ	kg m^{-3}	Luftdensitet
T	K	Temperatur
u	m s^{-1}	Vindkomponent i öst-västlig riktning
v	m s^{-1}	Vindkomponent i nord-sydlig riktning
w	m s^{-1}	Vertikal vindhastighet
f	s^{-1}	Coriolisparameter
g	m s^{-2}	Tyngdacceleration
V_g	m s^{-1}	Geostrofisk vind
θ	$^{\circ}$	Vindriktning
ΔV	m s^{-1}	Förändring i vindhastighet

Förkortningar

Förkortning	Beskrivning
ERA5	ECMWF Reanalysis version 5
GFS	Global Forecast System
LSTM	Long Short-Term Memory
NDBC	National Data Buoy Center
TKE	Turbulent Kinetic Energy
MSE	Mean Squared Error
MAE	Mean Absolute Error
CMAE	Circular Mean Absolute Error
BA	Balanced Accuracy

1 Inledning

I detta kapitel ges en beskrivning av projektet och varför det är relevant i dagens läge. Därefter följer syftet och målen med studien samt avgränsningar som gjorts i projektet.

1.1 Bakgrund

Segling är en komplex sport där det handlar till stor del om att förutse och anpassa sig till de yttre faktorerna ute på vattnet. Vindens styrka och riktning styrs av en mängd olika meteorologiska fenomen som är direkt avgörande hur strategiska och taktiska beslut fattas under tävlingsförhållanden. Om seglare kan förutse hur vinden kommer att bete sig under en dag kan fler korrekta beslut fattas.

Olympiska sommarspelen 2028 utspelar sig i Los Angeles och seglingstävlingarna kommer mer specifikt att ske utanför San Pedros kust. Under denna period förväntas väderfenomenet sjöbris spela en central roll för hur vinden kommer att bete sig ute på tävlingsområdet. Detta är särskilt relevant för seglare och tränare då en tillförlitlig analys av vindmönster ger en bättre bas för beslutstagande som i längden kommer att resultera i bättre tävlingsresultat.

Sjöbris är en termisk effekt som styrs av temperaturskillnader mellan hav och land och resulterar i en lokal vind [1]. När land värms upp under dagen stiger den varma luften och skapar ett lokalt lågtryck [1]. Den kalla luften från havet sveper in mot lågtrycket på land och detta skapar den vind som kallas för sjöbris. Sjöbrisens styrka och riktning kommer variera beroende på geografisk plats, lokal topografi och tid på året [1].

1.2 Syfte och mål

Syftet med studien är att bidra till ökad kunskap om hur lokala sjöbrisfenomen påverkar vindförhållanden i kustnära seglingsområden, samt hur dessa småskaliga meteorologiska processer kan analyseras och modelleras i prognossammanhang där hög rumslig och tidsmässig noggrannhet är avgörande.

Målet med projektet är att utveckla ett prognosverktyg med hjälp av maskininlärning som kan användas för följande punkter:

- Identifiera historiska fall av sjöbris baserat på meteorologiska data
- Generera timvisa prognoser för vindhastighet och vindriktning för nästkommande dag
- Uppskatta sannolikheten för förekomst av sjöbris under dagen

1.3 Uppgift

Sjöbris är ett vanligt förekommande fenomen under sommarhalvåret i Los Angeles. Statistiskt sett är det få dagar under juli månad utan sjöbris. Sjöbris varierar i styrka och riktning beroende på en mängd olika faktorer. Traditionella prognosmodeller har ofta begränsad förmåga att fånga lokala variationer i sjöbris, vilket kan leda till avvikelser

mellan prognostiserade och observerade vindförhållanden. Det faktum att modellerna inte heller är anpassade efter det specifika tävlingsområdet förstärker risken för avvikelser för tävlingsområdet. Huvudfokus för detta arbete är därför att skapa en daglig prognos för vindhastighet och vindriktning med maskininläring istället för att använda befintliga modeller.

Första steget för att nå projektets mål är att skapa en djupare förståelse av fenomenet sjöbris, vilket sker genom litteraturstudie. Att förstå vilka faktorer som har en inverkan på sjöbris vid tävlingsområdet kommer att bli avgörande för modellering och för att dra slutsatser kring resultatet.

Andra steget är att förstå vilken modell inom maskininläring som kan appliceras och att välja rätt modell. Tredje steget blir att hitta data som modellen kan träna på för att skapa en daglig prognos. Denna prognos kommer sedan att valideras med hjälp av lokal statistik och analyseras för att kalibreras och verifieras utifrån träffsäkerhet.

1.4 Avgränsningar

Projektet avgränsas till att studera sjöbrisens utveckling och påverkan på vindhastigheten och vindriktning nära markytan. Det geografiska området begränsas till kusten utanför Los Angeles, och mer specifikt området utanför San Pedros kust. Det geografiska området hålls begränsat för att möjliggöra en tillräckligt hög upplösning inom ramen för frågeställningen. Vidare avgränsas projektet till att analysera vindförhållandena mellan 10:00 och 16:00 lokal tid för maximalt fyra dagar under juli månad.

Python är det verktyg som primärt avses användas för att strukturera observationsdata, modellera och träna modellen. Inom ramen för projektet avses maskininlärningsmetoder undersökas med hjälp av PyTorch. I projektet avses modellen tillämpas retrospektivt med historiska och representativa väderförhållanden som indata för att identifiera typiska dygnsvariationer och övergripande mönster i sjöbrisens utveckling.

Projektet begränsas av att fysikaliska processer inte modelleras explicit, utan beräknas genom parametriserade statistiska samband. Även ett begränsat antal parametrar kommer att inkluderas i analysen av atmosfären. Identifieringen av sjöbris kommer att ske genom antagande i den tillgängliga data och inte genom observationer av sjöbriscirkulationen. Detta innebär att klassificeringen baseras på förutfattade antaganden. Valideringen av modellens resultat är begränsad till punktbaserad tillgänglig observationsdata från närliggande väderstationer. Ingen egen datainsamling genom fältmätningar ingår i projektet.

2 Teori

För att möjliggöra för utveckling av en tillförlitlig prognosmodell för sjöbris krävs en djupgående förståelse av de fysikaliska processer som styr fenomenet. Detta uppnås genom en litteraturstudie där relevanta meteorologiska teorier och modeller analyseras. Det specifika fokusområdet är på de faktorer som påverkar sjöbrisens utveckling i det aktuella geografiska området. Vidare behandlas de modellansatser som ligger till grund för prognostisering samt de verktyg som används i studien.

2.1 Sjöbris - översikt

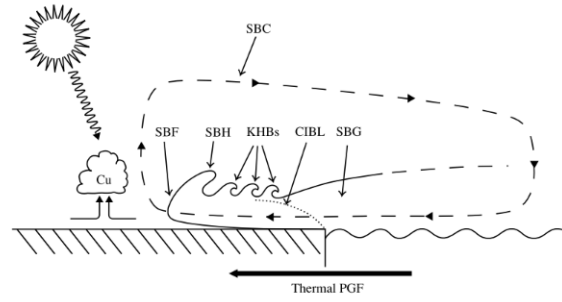
Sjöbris är en mesoskalig cirkulation som uppstår till följd av temperaturskillnader mellan land och hav [2]. Under dagen värms landytan upp snabbare än havet, vilket leder till att luften över land blir varmare och stiger [1]. Detta skapar ett lokalt lågtryck över land, vilket i sin tur driver ett inflöde av kallare luft från havet. Sjöbriscirkulationen förstärks successivt under förmiddagen och når sin fulla styrka under eftermiddagen [1]. I detta avsnitt beskrivs hur cirkulationen etableras, utvecklas över tid samt hur den kan representeras med förenklade modeller.

2.1.1 Sjöbrisens uppbyggnad

Sjöbrisen är en cirkulär process som drivs av temperaturskillnaden som skapas när luften över land värms upp snabbare än luften över havet [3]. På grund av lyftkraften som skapas av att luft med en högre temperatur har lägre densitet kommer lokala hög- och lågtryck att etableras. Detta skapar en tryckgradient som sedan driver sjöbriscirkulationen. I figur 1 beskrivs denna tryckgradient som en *Thermal PGF* [3]. Detta sker i gränsskiktet som är den nedersta delen av troposfären och utgör det område som direkt påverkas av förhållanden vid jordens yta [4]. I detta skikt dominerar turbulens, vilket ger osäkerheter för prognoser. Det är även i detta område som effekter av ytans termiska egenskaper visar sig. Den vind som upplevs kan beskrivas som en kombination av medelvinden och turbulensen och påverkas av faktorer som tryckskillnader, topografi och friktion [4]. Sjöbrisens uppbyggnad påverkas av de vertikala rörelserna i luften [4, 5]. Hur luften stiger och sjunker påverkas i sin tur av tryckskillnader, turbulens och luftens flytkraft [4].

Hela sjöbriscirkulationen beskrivs i figuren som SBC och utgör en vertikalt roterande mesoskalig cell med ett vinkelrätt flöde mot kusten nära markytan [3]. Även ett flöde av stigande luft inne vid land och ett sjunkande flöde ute över havet, samt ett återflöde, vilket skapar en sluten process. I figur 1 delas det marknära flödet in i de olika delarna:

- *Sea breeze gravity current* (SBG) som är det horisontella flödet av kall luft över havet.
- *Sea breeze front* (SBF) är fronten av sjöbrisen, där luften börjar stiga och hastigheten på det horisontella flödet minskar. I detta område sker det ofta stora förändringar i vind och temperatur. Här kan det byggas cumulusmoln.
- *Sea Breeze Head* (SBH) är den del bakom SBF, där den kalla luften påverkas av det vertikala flödet vilket skapar en topp i det horisontella flödet.
- *Kelvin-Helmholtz billows* (KHBs) är instabilitetsvågor i skjuvflöden som utvecklas av turbulens längs med den övre delen av SBG.
- *Convective internal boundary layer* (CIBL) är den ostabila regionen där den horisontella kallluften möter kusten.



Figur 1: Sjöbrisen är en cirkulär process [5]. Återgiven med tillstånd

2.1.2 Bjerknes cirkulations teori

Bjerknes cirkulationsteori beskriver med hjälp av en enkel matematisk modell hur en tryckgradient skapar en sjöbriscirkulation i en från början stilla atmosfär [3]. Teorin baseras på en förenklad matematisk modell där cirkulationen uttrycks som en linjeintegral.

$$\frac{D_a C_a}{D_t} = - \oint \frac{dP}{\rho} \quad (1)$$

Där $D_a C_a$ är materialderivatan med avseende på den aktuella luftmassan a , C_a är cirkulationen, P är trycket och ρ är densiteten för a . Med avseende på sjöbrisen kan ekvation 1 förenklas med hjälp av ideala gaslagen, vilket ger ekvation 2 [3]. Detta förenklar systemet till ett tryck nära landytan (p_0) och ett tryck i den övre delen av cirkulationen (p_1), men även medeltemperaturen i det vertikala luftflödet över land ($\overline{T_2}$) respektive hav ($\overline{T_1}$).

$$\frac{DC_a}{Dt} = R \ln\left(\frac{p_0}{p_1}\right) (\overline{T_2} - \overline{T_1}) \quad (2)$$

Där R är den ideala gaskonstanten. Denna modell kan skapa en prognos av vindhastigheten när sjöbrisen är fullt utvecklad [3]. Medelvindhastigheten ($\overline{U}$) kan uttryckas enligt ekvation 3.

$$\overline{U} = \frac{C_a}{2(H + L)} \quad (3)$$

Där H är höjden av området med cirkulerande luft och L är längden. Sätts ekvation 3 in i ekvation 2 fås följande [3]:

$$\frac{D\overline{U}}{Dt} = \frac{R \ln\left(\frac{p_0}{p_1}\right) (\overline{T_2} - \overline{T_1})}{2(H + L)} \quad (4)$$

Ekvation 4 beskriver hur vindhastigheten utvecklas över tid på grund av sjöbrisen [3]. Denna modell skapar en mycket förenklad bild av sjöbrisen. Resultat från dessa beräk-

ningar kommer därför inte att stämma överens med verkligheten, för en mer komplett modell borde flera faktorer inkluderas [3].

2.2 Tryckfält

Tryckgradienten (∇p) beskriver hur trycket förändras och utgör en central drivkraft bakom atmosfäriska rörelser [6]. Detta definieras enligt ekvation 5. $\frac{\partial p}{\partial x}$ representerar tryckets partiella derivata i x-led.

$$\nabla p = \left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}, \frac{\partial p}{\partial z} \right) \quad (5)$$

Tryckgradienten driver utvecklingen av sjöbris och kan beskrivas på olika sätt i samband med fenomenet [3]. Tre av dessa teorier är *upward theory*, *sideways theory* och *mixed theory*, men de är alla fysikaliskt inkonsekventa. Skillnaden mellan teorierna är hur de beskriver starten av cirkulationen, det vill säga, om det övre horisontella flödet eller det kalla flödet nära marken startar först. Det finns olika alternativ som beror på hur luftens expansion tolkas i sammanhanget [3].

Sjöbriscirkulationen påverkas inte bara av den lokala uppvärmningen utan också av ljudvågor [3]. Under utvecklingen av SBC sker det konstant en hydrostatisk anpassning [3]. Denna är svår att ta i beräkning, det finns därför icke-hydrostatiska modeller som bortser från den. En följd av den hydrostatiska anpassningen är att det bildas tryckvågor när luften värms adiabatiskt och expanderar. Tryckvågorna sprider sig sedan i alla riktningar och påverkar etableringen av tryckgradienten. De uppåtgående tryckvågorna skapar då en tryckökning i atmosfären ovanför den varma markytan medan de vågor som sprids horisontellt skapar ett tryckfall över land och en tryckökning över havet.

När cirkulation är etablerad kan fler lokala avvikelser mellan aktuellt tryck och bakgrundstrycket skapas, dessa avvikelser kallas tryckstörningar och uppstår vid sjöbrisfronten då kallluften möter det vertikala flödet [3]. Detta samtidigt som det bakom SBF, över land, sker en adiabatisk uppvärmning av luften samt att det sker en subsidens av den kallare luften i atmosfären. Detta påverkar tryckfältet då luften vid SBF har en uppåtgående riktning samtidigt som den kyls ner och får en högre densitet [3]. Samtidigt, bakom fronten, är det ett flöde med nedåtgående riktning som värms upp och har en lägre densitet. Därigenom bildas en tryckgradient med horisontell riktning som driver returströmmen högre upp i troposfären. Returströmmen är alltså inte endast en kompensation för flödet nära land utan drivs av en egen tryckgradient som uppkommer som en följd av advektion av värme [3]. Detta betyder att sjöbrisens styrka påverkas av hur temperaturfältet omfördelas i höjddled.

2.3 Bakgrundsvind

Sjöbriscirkulationen startar när vinden som uppstår av den mesoskaliga tryckgradienten adderas till den redan befintliga bakgrundsvinden nära marken [3]. Bakgrundsvinden är den vind som uppstår i atmosfären på grund av tryckskillnader i vädersystem på en större skala. Med andra ord kan bakgrundsvinden beskrivas som vinden utan de lokala

effekterna. Bakgrundsvindens styrka och riktning har stor påverkan på sjöbrisens form och dess position relativt kusten [3].

2.3.1 Sjöbrisindex

Sjöbrisindex är ett index som enkelt beskriver om sjöbris kan överta bakgrundsvinden [3]. Den jämför bakgrundsvinden med temperaturskillnaden mellan hav och land. Detta genom ett förhållande mellan den kinetiska energin i bakgrundsvinden och den termiska drivningen i temperaturskillnaden. Sjöbrisindex (ϵ) beräknas med ekvation 6 [3].

$$\epsilon = \frac{|U|^2}{C_p \Delta T} \quad (6)$$

Där U representerar vindhastigheten, C_p den specifika värmekapaciteten i luften och ΔT temperaturskillnaden mellan hav och land. Ett stort värde på ϵ ger ingen sjöbris då den undertrycks av bakgrundsvinden eller när temperaturskillnaden inte är tillräckligt stor [3]. Det kritiska värdet ändras beroende på de geografiska egenskaperna som finns i området [3].

2.3.2 Olika typer av sjöbris

Sjöbris kan delas in i tre huvudsakliga kategorier: ren, korkskruv och bakdörr [3]. Vilken typ av sjöbris som uppstår beror på aktuell bakgrundsvind. På en östkust i den norra hemisfären kan de olika kategorierna beskrivas enligt följande [3].

Ren sjöbris uppstår vid väldigt lätt bakgrundsvind, vilket resulterar i att den har liten påverkan på sjöbrisens uppbyggnad och egenskaper [3]. Ren sjöbris ses som det mest idealiserade fallet eftersom den generella beskrivningen av sjöbris stämmer väl in på denna kategori [3]. Det som karaktäriserar en ren sjöbris är:

- En tydlig tryckgradient mellan land och hav
- En vertikal, i huvudsak sluten, cirkulationscell
- En period av lugna förhållanden med svaga eller lugna vindar innan sjöbris etableras, då den synoptiska tryckgradienten tillfälligt balanserar den lokala termiska drivningen
- När långt in över land, upp mot 130 km.

Korkskruvsjöbris sker vid befintlig bakgrundsvind och har både en komponent längs med kusten och en mot kusten [3]. Enligt Buys-Ballots lag innebär detta att luft nära marken sprids bort från kustområdet, vilket leder till att mängden luft över området minskar. Detta kallas horisontell divergens [3]. Divergensen är en följd av friktion och corioliskraften och gör att luft från högre nivåer sjunker ner mot markytan. Vilket i sin tur förstärker initieringen av sjöbris. Riktningen på bakgrundsvinden gör att den inte är parallell med den termiska tryckgradienten som driver sjöbris [3]. Detta medför att den inte behöver övervinna den synoptiska gradienten.

Det som karaktäriserar korkskruvsjöbris är [3]:

- Uppstår fortare än ren då sjöbriscirkulationen förstärks av divergens
- Avsaknaden av en tydlig lugn period innan etablering
- Skruvformad, helikal, cirkulationscell i norrgående riktning
- Gradvis vridning av vinden i moturs riktning från nordväst till nordöst
- Möjlighet att nå kusten även vid svagare termisk tryckgradient jämfört med ren sjöbris

Bakdörrsjöbris uppstår, likt korkskruvsjöbris, när bakgrundsvinden har parallella och vinkelräta komponenter mot kusten [3]. Till skillnad från korkskruvsjöbris är den parallella komponenten riktad söderut för en östkust i den norra hemisfären [3].

Det som kännetecknar bakdörrsjöbris är [3]:

- Svagare i intensitet jämfört med ren och korkskruv
- Etablering sker senare under dagen
- Etableringen är pulserande snarare än kontinuerlig
- Gradvis vridning av vinden i medurs riktning från nordväst till nordöst
- Skruvformad cirkulationscell i södergående riktning

Bakdörrsjöbris kan bara nå kusten vid en starkare termisk tryckgradient jämfört med de andra sjöbriskategorierna [3].

2.4 Jordens rotation

När rörelser analyseras i atmosfären görs det med ett koordinatsystem som har samma rotation som jorden. Detta innebär att systemet är icke-inertialt, det vill säga koordinatsystemet accelererar, vilket medför att Newtons rörelselagar inte kan tillämpas direkt [6]. Därför måste Newtons rörelselagar modifieras med hjälp av fiktiva krafter, till exempel corioliskraften och centrifugalkraften. Dessa skenbara krafter tas med i beräkningen när observationer i roterande system beräknas, så att Newtons ekvationer kan tillämpas. En av de centrala krafterna när vind studeras är Corioliskraften [4, 6].

2.4.1 Corioliskraften

Corioliskraften uppstår när ett objekt sätts i rörelse i ett roterande system [6]. Kraften verkar vinkelrätt mot rörelseriktningen och påverkar därför objektets bana utan att påverka dess hastighet. Den matematiska formeln för Corioliskraften F_c beskrivs matematiskt enligt ekvation 7 [6].

$$F_c = -2\Omega \times \mathbf{V} \quad (7)$$

Där Ω representerar jordens rotation och V är hastighetsvektorn för objektets rörelse..

I meteorologiska samband används oftast en förenklad form av ekvationen som kallas för Coriolisparametern f som beräknas genom ekvation 8 [6].

$$f = 2\Omega \sin\phi \quad (8)$$

ϕ är latituden och denna parameter styr Coriolisparameterns styrka. På ekvatorn är $f = 0$ och effekten blir därmed mycket liten, men effekten ökar mot polerna på jorden där sinusvärdet ökar [6].

2.4.2 Corioliskraftens påverkan på vinden

Corioliskraften påverkar objekt i olika riktningar beroende på vilken sida av ekvatorn som objektet färdas på [6]. På det norra halvklotet drar objekt åt höger när de sätts i rörelse och på det södra halvklotet drar objekt åt vänster när de sätts i rörelse. Hur mycket en vind påverkas av Corioliskraften beror på vindhastigheten och på vilken latitud vinden befinner sig på [6].

När luften sätts i rörelse genom tryckgradienten kommer Corioliskraften påverka riktningen av luftpartiklarna fram till geostatisk balans mellan Corioliskraften och tryckgradienten uppstår [6]. I detta är Corioliskraften lika stark som tryckgradienten. Tryckgradienten drar vind från högtryck till lågtryck, medan Corioliskraften vill dra vinden åt höger på norra halvklotet. Geostatisk balans innebär i detta fall att vindriktningen till slut kommer att tangera isobarerna, istället för att blåsa från hög till lågt tryck.

Den geostrofiska vinden (V_g) beräknas genom ekvation 9 [6].

$$f\mathbf{k} \times V_g = \frac{-1}{\rho} \nabla p \quad (9)$$

V_g är den geostrofiska vindvektorn, ρ är luftens densitet och ∇p är tryckgradienten. Vid omskrivning av ekvation 9 till komponentform, fås 10 [6].

$$\begin{cases} u_g = \frac{1}{f} \frac{\partial p}{\partial y} \\ v_g = \frac{1}{f} \frac{\partial p}{\partial x} \end{cases} \quad (10)$$

u_g och v_g är de geostrofiska vindvektorerna i öst-väst- respektive nord-syd-riktning. Sambandet styrs enbart av coriolisparametern och tryckgradienten, men detta gäller enbart i den fria atmosfären vid perfekta förhållanden [6].

Mesoskaliga system, som sjöbris, uppnår sällan geostrofisk balans på grund av friktion från turbulens i gränsskiktet [4]. Corioliskraften påverkar därför till en viss grad sjöbrisens vindriktning, men även andra faktorer styr sjöbriscirkulationen [4].

2.5 Vertikala lufttransporter i gränsskiktet

Den vertikala lufttransporten i atmosfären sker när det är skillnader i atmosfärens stabilitet, vilket innebär att varm luft kan stiga och skapa vertikala rörelser [4]. När luften värms upp över land minskar dess densitet och stiger [4]. Varm luft har lägre densitet än kall och denna skillnad skapar en vertikal flytkraft, som i sin tur skapar konvektion i gränsskiktet. Detta är drivkraften till vertikal transport av värme och rörelsemängd.

Flytkraften beskrivs enligt ekvation 11 [4].

$$B = g \frac{\theta'}{\theta} \quad (11)$$

där g är gravitationen och θ är den potentiella temperaturen.

Ett mått på hur stabil atmosfären är i gränsskiktet är Brunt-Väisäläfrekvensen som visar sambandet mellan den vertikala gradienten av potentiell temperatur och beskrivs enligt ekvation 12 [4].

$$N^2 = \frac{g}{\theta} \frac{\partial \theta}{\partial z} \quad (12)$$

När $N^2 > 0$ så är atmosfären i gränsskiktet stabil, vilket betyder att det är inga eller små vertikala transporter i gränsskiktet [4]. När $N^2 < 0$ ses atmosfären ha instabila skiktningar där konvektion och turbulens kan uppstå.

2.6 Turbulens i atmosfäriska gränsskiktet

Turbulensen i det atmosfäriska gränsskiktet har en central roll i sjöbriscirkulationen [4]. När luftflödet bryts upp i oregelbundna virvlar, så kallade *eddies*, skapas turbulens mellan olika luftmassor. Genom turbulens transporteras rörelsemängd, värme och fukt vertikalt mellan marken och den övre delen av gränsskiktet [4, 6]. Den viktigaste orsaken till turbulens är temperaturdifferens som uppstår när solens strålar värmer upp land [4].

2.6.1 Reynoldsdekomposition av vindhastigheten

När turbulens analyseras delas de atmosfäriska flödets variabler upp i en medelkomponent och en fluktuation [4]. Denna metoden kallas Reynoldsdekomposition och för vindhastigheten kan den skrivas som, se ekvation 13.

$$u = \bar{u} + u' \quad (13)$$

$\bar{u}$ representerar medelvärdet av hastigheten u och u' representerar den turbulenta fluktuationen kring medelvärdet. Reynoldsdekomposition kan användas för att göra samma uppdelning på vertikal vindhastighet och temperatur [4]. När rörelsemängd och transport av energi sker på grund av turbulenta rörelser i den fria atmosfären är det möjligt att beskriva sambanden med olika Reynoldsdekompositioner.

2.6.2 Turbulent transport av rörelsemängd

Den turbulenta transporten av rörelsemängd kan beskrivas med Reynolds-spänningar och det vertikala momentumet kan uttryckas enligt ekvation 14 [4].

$$\tau = -\rho \overline{u'w'} \quad (14)$$

ρ är luftens densitet, u' är den turbulenta fluktuationen i horisontell vindhastighet och w' är den turbulenta fluktuationen i vertikal vindriktning. Uttrycket visar hur mycket turbulensen transporterar vindens horisontella rörelsemängd vertikalt i atmosfären [4, 6]. I gränsskiktet är vinden starkare på högre höjd än på marken på grund av friktion [4]. Turbulenta virvlar blandar en starkare vind på högre höjd med den svagare vinden på marken vilket leder till den vertikala transporten av vindens rörelsemängd [4].

2.6.3 Turbulent transport av värme

Transport av värme sker med den turbulenta värmeffluxen och beskrivs enligt ekvation 15 [4].

$$\overline{w'T'} \quad (15)$$

w' är den turbulenta fluktuationen i vertikal vindriktning och T' är den turbulenta fluktuationen i temperaturen. Uttrycket beskriver hur turbulenta rörelser flyttar varm och kall luft vertikalt i atmosfären [4]. Värmeffluxen påverkar hur effektivt markvärmen transporteras upp i atmosfären och därmed utvecklingen av konvektiva rörelser.

En svårighet med de turbulenta virvlarna är att de är mycket små, ibland bara några centimeter [4]. Många modeller har inte så små gridceller och därför kan inte modellerna simulera varje virvel. Istället parametreras varje turbulent virvel, vilket innebär att ekvationen skrivs om till en enklare matematisk relation. Parametriseringen för värmetransport beskrivs ofta enligt ekvation 16 [4].

$$\overline{w'T'} = -K_h \frac{\partial T}{\partial z} \quad (16)$$

K_h är en turbulent diffusivitetskoefficient och beskriver hur stark turbulensen är och koefficienten beror på atmosfärisk stabilitet och turbulensenergi. $\frac{\partial T}{\partial z}$ är den vertikala temperaturgradienten. Ekvationen beskriver hur turbulenta rörelser ger upphov till en effektiv diffusion av värme från varmare områden till kallare [4].

2.6.4 Turbulent kinetisk energi

Turbulent kinetisk energi produceras genom vindskjuvning och lyftkrafter inom instabila atmosfäriska skikt [4]. Vindskjuvning uppstår när det finns skillnad i vindhastighet på olika höjder och sker mellan olika luftlager, som i sin tur kan resultera i turbulenta virvlar. Energi som genereras genom vindskjuvning och flytkraft tillförs främst de större

turbulenta virvlarna [4, 7]. Denna energi överförs därefter genom en energikaskad från större till mindre skalor, där allt mindre virvlar dominerar [7]. Då börjar den molekylära viskositeten dominera och den turbulenta kinetiska energin omvandlas till värme genom friktion. Processen kallas dissipation och är slutet på energiöverföringar inom turbulenta flöden [4, 7].

Turbulent kinetisk energi är summan av kinetisk energi relaterat till de turbulenta hastighetsfluktuationerna [4].

$$TKE = \frac{1}{2}(\overline{u'^2} + \overline{v'^2} + \overline{w'^2}) \quad (17)$$

där u' är den turbulenta fluktuationen i vindhastighet i öst-väst-riktning, v' är den turbulenta fluktuationen i vindhastighet i nord-sydlig-riktning och w' är den turbulenta fluktuationen i vertikal vindhastighet.

2.6.5 Turbulens i sjöbriscirkulation

Turbulens har en central roll i sjöbriscirkulationen och framför allt när det kommer till sjöbrisfronten. När den kalla luften från havet blandas med den varma luften från land skapas vertikala rörelser och instabil skiktning som skapar turbulens. Blandningen av luftmassorna påverkar temperaturen i gränsskiktet och påverkar hur snabbt sjöbrisfronten rör sig in över land [4, 3, 8].

2.7 Modeller

Studie av sjöbrisen kan delas in i två kategorier, analytiska och numeriska.

2.7.1 Analytiska modeller

Inom meteorologi används oftast analytiska modeller för att beskriva atmosfäriska processer med hjälp av enklare matematiska relationer [6]. Modellerna bygger i de flesta fall på idealiserade antaganden. Analytiska modeller används i första hand för utveckling av teoretisk förståelse som senare används i numeriska modeller av atmosfäriska cirkulationen [6].

När sjöbris studeras med analytiska modeller är tryckgradientkraften en vanlig förekommande ekvation [6]. Den används för att beskriva hur temperaturen mellan hav och land skapar en tryckgradient [3]. Tryckgradientkraften beskrivs enligt ekvation 18 [6].

$$F_{pgf} = -\frac{1}{\rho}\nabla p \quad (18)$$

ρ är luftens densitet och ∇p representerar tryckgradienten i atmosfären. Relationen visar att luft accelererar från högt till lågt tryck. Denna ekvation tillsammans med enkla antaganden om temperaturkontrasten mellan hav och land används för att beskriva hur sjöbriscirkulationen uppstår [3, 8].

Temperaturgradienten studeras också i analytiska modeller genom tryckfält som påverkas av den hydrostatiska balansen [6]. När temperaturen varierar horisontellt förändras tryckfördelningen på grund av att när varm luft expanderar blir den mer spridd, vilket gör att luftpelare över det varma området blir högre [6, 3]. När trycket i atmosfären förändras skapas en horisontell tryckgradient mellan olika områden [6] och denna tryckgradient är en drivande faktor bakom sjöbrisen [3, 9]. Ett uttryck som ofta används i analytiska modeller för att koppla samman tryckgradienten och temperaturgradienten och approximeras enligt ekvation 19 [6].

$$\frac{\partial p}{\partial x} \approx -\frac{p}{T} \frac{\partial T}{\partial x} \quad (19)$$

p är trycket, T är temperaturen och $\frac{\partial T}{\partial x}$ presenterar den horisontella temperaturgradienten. Ekvationen visar hur differenser i temperatur i horisontell riktning kan generera en motsvarande tryckgradient [6].

Analytiska modeller ger en konceptuell förståelse för hur sjöbrisen styrs med enkla matematiska koncept [9]. Modellerna är idealiserade och tar inte hänsyn till topografi, bakgrundsvind eller turbulenta processer. För att använda modeller som är mindre idealiserade krävs mer avancerade numeriska modeller som kan lösa atmosfärens rörelseekvationer.

2.7.2 Numeriska modeller

Numeriska modeller använder lösningar av fysikaliska ekvationer för att lösa luftens rörelse och energibalans [6]. Atmosfärens dynamik styrs till grunden av fysikaliska lagar som uttrycks i partiella differentialekvationer [6]. Ekvationerna innefattar variabler för vindhastighet, temperatur, tryck och densitet som förändras över tid och rum. Det är ofta svårt att lösa den typen av komplexa ekvationer för realistiska atmosfäriska förhållanden. Därför används numeriska metoder där ekvationerna löses approximativt över små gridceller [6].

Atmosfären delas upp i ett tredimensionellt rutnät där styrande ekvationerna beräknas i varje punkt [6]. Utvecklingen beräknas stegvis genom små tidssteg framåt, vilket möjliggör prognoser av atmosfärens tillstånd. De viktigaste ekvationerna i numeriska modeller utgörs av rörelse-, kontinuitets-, den termodynamiska energiekvationen och tillståndsekvationen för en ideal gas [6]. Rörelseekvationerna kan i förenklad form uttryckas enligt ekvation 20 [6].

$$\frac{D\mathbf{V}}{t} = -\frac{1}{\rho} \nabla p + \mathbf{g} - 2\boldsymbol{\Omega} \times \mathbf{V} \quad (20)$$

$\mathbf{V}$ är vindhastigheten, p är trycket, ρ är luftens densitet, $\mathbf{g}$ är gravitationen och $2\boldsymbol{\Omega} \times \mathbf{V}$ representerar corioliskraften. Ekvationen beskriver hur luften påverkas av tryckgradients, gravitation och corioliskraften.

Med kontinuitetsekvationen kan massbevarandet beskrivas genom ekvation 21 [6].

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0 \quad (21)$$

ρ är luftens densitet och $\mathbf{V}$ är vindhastighetsvektorn. Den första termen $\frac{\partial \rho}{\partial t}$ beskriver hur densiteten ändras över tid i en viss punkt i atmosfären. Den andra termen $\nabla \cdot (\rho \mathbf{V})$ är divergensen av massflödet, vilket beskriver nettotransporten av luft som transporteras in eller ut ur ett område. Ekvationen visar att massa inte försvinner utan bara transporteras genom atmosfären.

Temperatures utveckling beskrivs med den termodynamiska energiekvationen som beskrivs i ekvation 22 [6].

$$\frac{DT}{Dt} = \frac{Q}{c_p} \quad (22)$$

T är temperaturen och Q är värmetillskott och c_p är luftens specifika värmekapacitet. Vänstra sidan av ekvationen beskriver hur temperaturen förändras för luft som rör sig genom atmosfären. På högra sidan påverkas temperaturförändringen av olika energikällor som strålning och latent värme. Ekvationen visar hur till- eller bortförsel av energi påverkar temperaturen på luften.

Sambandet mellan tryck, temperatur och densitet i atmosfären uttrycks med tillståndsekvationen för en ideal gas, vilken ges av ekvation 23 [6].

$$p = \rho RT \quad (23)$$

där p är trycket, ρ är luftens densitet, T är temperatur och R är gaskonstanten för luft.

2.8 Maskininlärning

Maskininlärning (ML) är en underkategori inom artificiell intelligens (AI) med fokus på utveckling och användning av algoritmer som kan "lära" sig olika mönster i träningsdata för att sedan dra korrekta slutsatser om nya data [10]. Huvudsyftet med ML är att optimera modellers prestanda på ett dataset som liknar de verkliga problemen som modellen ska användas för [10].

Till skillnad från traditionell programmering, där regler definieras i förväg, bygger ML på att modellen tränas på så kallad träningsdata [10]. Genom träningsprocessen lär sig modellen att identifiera statistiska samband i datan, vilket möjliggör för modellen att ta beslut utan explicita instruktioner. En central princip inom maskininlärning är modellens generaliseringsförmåga, vilket innebär att modellen inte enbart ska prestera på träningsdatan, utan även kunna producera tillförlitliga resultat vid verkliga tillämpningar. Därmed är slutmålet med ML att modellen ska kunna prediktera eller klassificera korrekt vid analys av nya dataset.

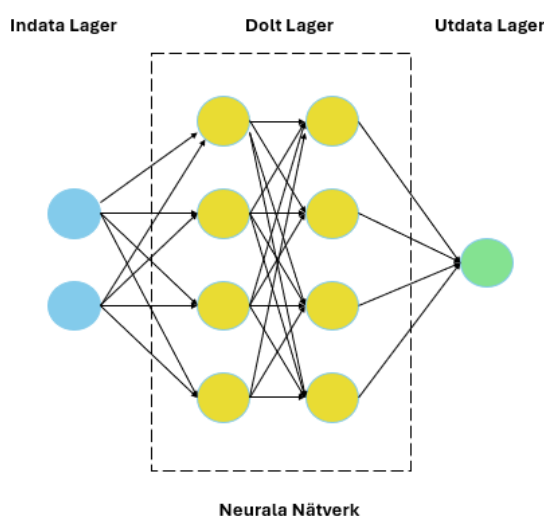
All maskininlärning kan delas in i en av tre kategorier: supervised learning, unsupervised learning eller reinforcement learning [10]. Vilken av kategorierna ML delas in i beror på målet med träningen och i många fall även utifrån vilken typ av träningsdata som används.

- Supervised learning: modellen tränas för att förutspå den korrekta outputen för en given input. Denna kategori gäller för uppgifter som kräver en viss grad av noggrannhet relativt till en utomstående sanning, kallad ground truth, såsom klassificering eller regression.
- Unsupervised learning: modellen tränas att skilja på inbyggda mönster, kopplingar och villkor i data. Träningsdatan för unsupervised learning använder inte uppgifter som involverar utomstående ground truth som outputen ska jämföras med.
- Reinforcement learning (RL): modellen tränas för att utvärdera sin miljö och ta beslut som ger störst “belöning”. RL-scenarier baseras inte på ett singulärt rätt svar, utan på “bra”, “dåliga” eller neutrala beslut.

Det finns även andra hybridvarianter för ML, exempelvis semi-supervised learning, vilket kan ge modellen en bättre optimering än de tre kategorierna nämnda ovan [10].

2.8.1 Deep Learning

Deep learning är en specialgren inom maskininlärning som använder artificiella neurala nätverk med många lager istället för definierade algoritmer som används i traditionell ML [10]. Neurala nätverk är löst inspirerade av den mänskliga hjärnan och består av sammankopplade lager av “neuroner” (även kallade noder), där varje nod utför sin egen matematiska operation (kallad aktiveringsfunktion). Utdata från varje nods aktiveringsfunktion fungerar som input till noderna i nästa lager och så vidare fram till sista lagret där den slutgiltiga outputen beräknas. Varje nods aktiveringsfunktion är icke-linjär, vilket möjliggör för deep learning modellen att hantera komplexa mönster och villkor, se figur 2.



Figur 2: Hur neurala nätverk är uppbyggda

2.8.2 PyTorch

PyTorch är ett ramverk baserat på öppen källkod som används inom deep learning för att bygga neurala nätverk och träna modeller genom att kombinera ML biblioteket Torch med en Pythonbaserad API [11]. Ramverket stödjer många olika neurala nätverksstrukturer, från enkla linjära regressioner till komplexa invecklade neurala nätverk. PyTorch är byggt på programmeringsspråket Python och erbjuder ett omfattande bibliotek med redan konfigurerade och förtränade modeller som har tränats på omfattande dataset inom exempelvis bildanalys, språkbehandling och ljudbearbetning. Detta minimerar både tid och arbete för kodning av sofistikerade deep learning modeller och möjliggör partiell körning av koden, vilket gör PyTorch utmärkt för prototyper och snabb utveckling av modeller.

Kärnan i PyTorch utgörs av flerdimensionella matriser, kallade tensorer, där data och modellparametrar lagras och bearbetas med stöd av GPU för snabba beräkningar [11]. Tensorerna kan även lagra icke numerisk information, såsom ljud och bild genom att konvertera dessa numeriskt. Förutom att koda modellens inputs och outputs, kodar tensorerna även modellparametrar såsom vikt, antaganden och gradienter, vilket lärts i maskininlärning. Denna egenskap möjliggör automatisk beräkning av derivator genom autograd modulen, vilket är ett av PyTorchs viktigaste egenskaper [11]. Autograd använder den matematiska kedjeregeln för sin automatiska urskiljning. Modulen beräknar komplexa derivator genom att dela upp dem i enklare derivator och kombinerar dem senare. Detta möjliggör optimering av modellens parametrar genom bakåtpropagering, där felet mellan modellens prediktion och det faktiska utfallet används för att justera vikterna. Autograd beräknar och sparar automatiskt gradienterna för alla matematiska operationer som utförts i en dynamisk beräkningsgraf. Detta innebär att modellen byggs upp och utförs stegvis under körning och därmed ger stor flexibilitet och underlättar utveckling, testning och felsökning [11].

Utöver tensorer, autograd och dynamiska beräkningsgrafer har PyTorch bland annat följande grundläggande komponenter:

- Neurala Nätverks API: Ett modulärt ramverk för att bygga neurala nätverk med fördefinierade lager, aktiveringsfunktioner och förlustfunktioner [12].
- DataLoaders: Verktyg för att effektivt hantera data. Erbjuder batching, shuffling och parallell uppladdning av data. Tar bort komplexiteten med förprocessering av data och iterering, vilket möjliggör för optimerade träningsloopar [12].

2.9 Dataset

De olika dataseten för väder- och vinddata som används i projektet är ERA5 och GFS. I denna sektion förklaras de olika typerna av data samt hur respektive insamling sker.

2.9.1 ERA5

ERA5 är den femte generationen av ECMWF:s reanalysverktyg som förser användaren med historiska väderdata från 1940 och framåt [13]. Reanalys kombinerar modelldata med observationer runtom världen för att skapa ett komplett och konsekvent dataset

genom att använda fysiklagar. Denna metod, som kallas data assimilation, är baserad på numeriska väderprognoscenter, där efter ett bestämt antal timmar kombineras en tidigare prognos med nya observationer [13]. Kombinationen sker på ett optimalt sätt för att producera en ny bästa uppskattning av atmosfärens tillstånd, vilket kallas analys. Från detta kan en ny uppdaterad och förbättrad prognos skapas. Reanalys fungerar på samma sätt där skillnaden är att upplösningen är lägre, vilket möjliggör för dataset från flera årtionden tillbaka [13]. Reanalys erbjuder mer tid för att samla in observationer och tillåter förbättrade observationer att integreras med historiska data, vilket i slutändan är till fördel för kvaliteten på reanalysen [13].

ERA5 erbjuder timvisa uppskattningar för ett stort antal atmosfäriska, havs-våg och land-yta datapunkter [13]. En uppskattning av osäkerheten samplas var tredje timme. Uppskattningen av osäkerheten är närbesläktad med de tillgängliga observationssystemen, som har utvecklats över tiden. Dessa indikerar också områden som är känsliga för flöden.

De tillgängliga dataseten är rutnätade på ett 0.25° latitud-longitudrutnät för reanalys och 0.5° för uppskattningen av osäkerheten [13]. Vid nedladdning av ERA5 timvis data väljs de datapunkterna som ska studeras utifrån någon av de 14 tillgängliga huvudområdena, vilka visas i tabell 1. Minst ett val av datapunkt måste väljas. Data som laddas ner har filformatet GRIB.

Tabell 1: Huvudområden för ERA5 väderdata [13]

Temperature & Pressure	Precipitation and Rain
Wind	Snow
Mean Rates	Soil
Radiation and Heat	Vertical Integrals
Clouds	Vegetation
Lakes	Ocean Waves
Evaporation and Runoff	Other

2.9.2 GFS

Global Forecast System, GFS, är en väderprognosmodell från National Centers for Environmental Prediction, NCEP, som genererar data för ett dussintal atmosfäriska och land-markvariabler, vilket inkluderar temperaturer, vind, nederbörd, fuktighet i jorden och atmosfärisk ozonkoncentration [14]. GFS parar ihop fyra separata modeller, atmosfär, havsmodeller, land/mark modeller och sjöis, som sedan arbetar tillsammans för att skildra väderförhållandet på ett korrekt sätt.

GFS används för att ge globala atmosfäriska prognoser vid 13 km upplösning och används som initiala och/eller gränsvillkor för andra NCEP prognosystem [14]. De initiala tillstånden för GFS prognosen ges av Global Data Assimilation System, GDAS. GDAS kör ett fyrdimensionellt hybrid ensamble-variational data assimileringssystem som producerar de initiala tillstånden för GFS prognosen.

GFS rutnät består av ett $0.25^\circ \times 0.25^\circ$ globalt latitud-longitud rutnät [15]. GFS körs dagligen vid fyra tillfällen och kan producera prognoser upp till 16 dagar i förväg [16].

Prognoskomponenten använder en kubisk finit volymmodell, även kallad FV3, med en upplösning på ungefär 13 km och där atmosfären är kopplad till NCEPs globala vågmodell [16]. I det vertikala är modellen indelad i 127 vertikala lager och producerar en timvis prognos för de första 120 timmarna och sedan var tredje timme för dag 5 till 16.

De olika variablerna som tillhandahålls av GFS är listade i tabell 2

Air temperature	Albedo	Cloud frequency	Cloud liquid water/ice
Dew point temperature	Evaporation	Geopotential height	Gravity wave
Heat flux	Humidity	Hydrostatic pressure	Ice depth/thickness
Ice extent	Land-use/land-cover classification	Longwave radiation	Maximum/minimum temperature
Planetary boundary layer	Potential temperature	Precipitation amount	Precipitation rate
Runoff	Sea-level pressure	Sea-surface temperature	Shortwave radiation
Skin temperature	Snow depth	Snow-water equivalent	Soil moisture/water content
Soil temperature	Surface winds	Total precipitable water	Upper-air temperature
Upper-level winds	Vertical wind velocity/speed	Visibility	Vorticity
Wind shear			

Tabell 2: Urval av meteorologiska variabler som kan hämtas från GFS [14]

2.10 Observationssystem och väderbojar

Utvecklingen av observationssystem har skett under senare delar av 1900-talet samt under 2000-talet i takt med den tekniska utvecklingen av internet [17]. Inledningsvis användes manuell dokumentation via fartyg och olika kuststationer, vilket gjorde att datatäckningen blev något begränsad.

Utvecklingen av observationssystem har varit viktig för både kvaliteten och jämförbarheten i de meteorologiska datorerna [17]. Den första publiceringen av *Survey on the Use of the Guide to Instruments and Methods of Observation (WMO-No.8)*, gjordes på 1950-talet efter detta har det ställts högre krav på observationssystem. Övergången från manuell till en mer automatiserad process av observationssystem har lett till att väderbojar i stora observationsnätverk utgör en stor del av den meteorologiska och oceaniska data som existerar idag [17]

Väderdata används till stora delar i marin observation och har som primär uppgift att samla in meteorologiska och oceaniska data [17]. Enligt *Survey on the Use of the Guide to Instruments and Methods of Observation (WMO-No.8)* är det viktigt med ett standardiserat sätt att observera data för att kunna jämföra meteorologiska mätningar globalt med prognoser och forskning. Variabler som är standardiserade och mäts av väderbojar är exempelvis vindhastighet, vindriktning, lufttemperatur, havstemperatur, lufttryck och luftfuktighet [18].

2.10.1 Väderbojar utanför San Pedro

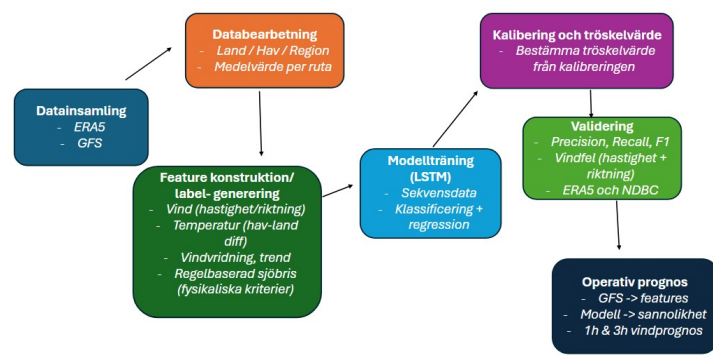
Längs Kaliforniens kust finns flertalet väderstationer som hanteras av NOAA NDBC, National Oceanic and Atmospheric Administration i USA [19]. Data importeras från väderstationerna och är tillgänglig för allmänheten. I anslutning till hamnen utanför San Pedro finns det en väderstation som heter Los Angeles, CA-NOAA NDBC station 9410647 [19]. Stationen producerar timupplöst vindhastighet, vindriktning, lufttryck och temperatur. All data finns tillgänglig för nedladdning från 2010 till idag och kommer att användas som representation för mer kustnära vindförhållanden.

3 Metod

I detta kapitel beskrivs den metodik som användes för att utveckla och utvärdera en datadriven modell för prognostisering av sjöbris. Metoden omfattar en komplett kedja från datainsamling till operativ prognos för tävlingssegling och är utformad för att vara reproducerbar. Metoden omfattar följande fyra huvuddelar.

- Hämtning och bearbetning av data
- Konstruktion av fysikaliskt motiverade indata och målvariabler
- Träning av en sekvensbaserad LSTM-modell (Long Short-Term Memory) och validering.
- Applicering av den tränade modellen på GFS-prognoser för att skapa en prognos för sjöbris för nästkommande dag

För en tydligare överblick av metoden visas arbetsflödet i flödesschemat i figur 3. Den beskriver modellkedjan från insamling och bearbetning av data, med konstruktion av egenskaper och etiketter, samt LSTM-träning, till en operativ prognos applicerbar på en GFS-prognos.



Figur 3: Flödesschemat visar modellkedjan från datainsamling till en operativ prognos applicerbar på en GFS-prognos

3.1 Databearbetning

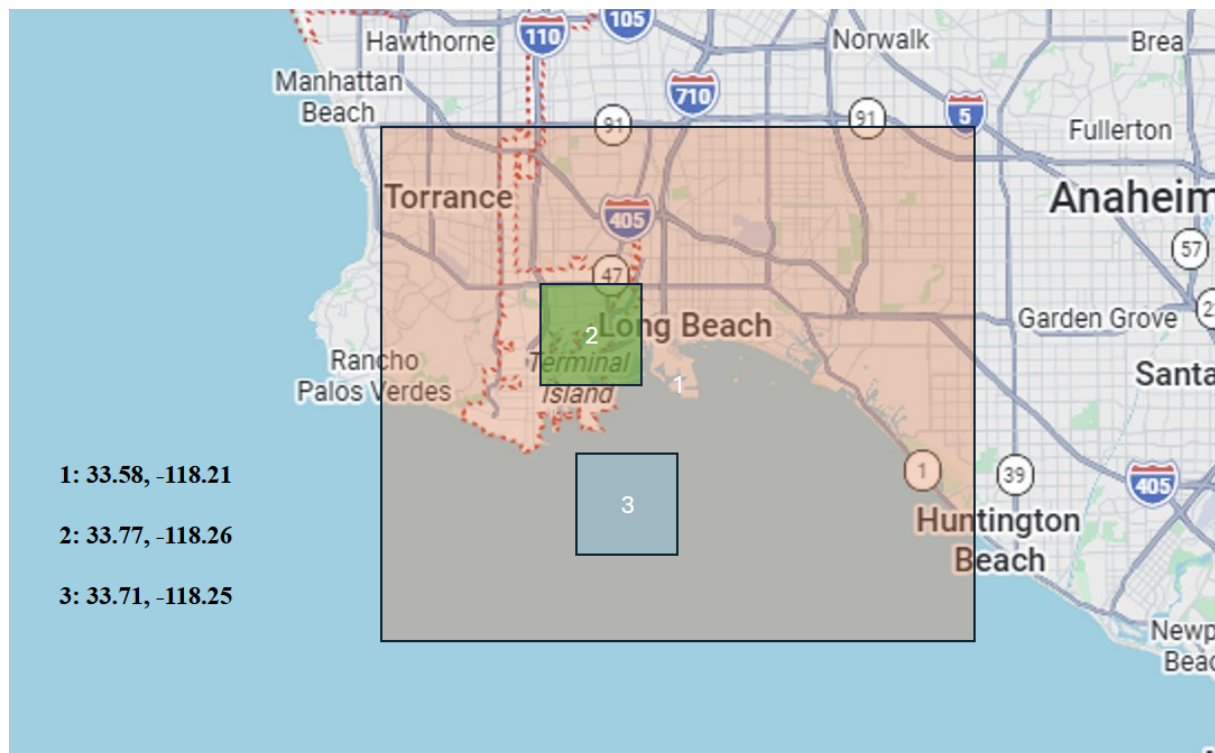
Meteorologiska data hämtades från ERA5 (reanalysdata) och GFS (prognosdata). De nedladdade datamängderna omfattade relevanta variabler såsom vindhastighet, vindriktning, temperatur och tryck för det geografiska området utanför San Pedros kust.

Studieområdet delades upp i tre geografiska rutor: havsruta, landruta och regionruta. Detta genomfördes för att fånga de termiska kontrasterna som driver sjöbris. Koordinaterna som användes för varje ruta listades nedan och 4 visualiserade detta i förhållande till det verkliga området.

- Havsruta (SEABOX): med området minimum latitud 33.52, maximum latitud 33.64,

minimum longitud -118.30 , maximum longitud -118.12 .

- Landruta (LANDBOX): med området minimum latitud 33.70 , maximum latitud 33.84 , minimum longitud -118.34 , maximum longitud -118.18 .
- Regionsruta (REGIONBOX): med området minimum latitud 33.50 , maximum latitud 33.92 , minimum longitud -118.55 , maximum longitud -117.95 .



Figur 4: Boxarna som illustreras är havsrutan, landrutan och regionrutan

För att filtrera bort och reducera lokalt brus användes areamedelvärden för ett antal relevanta variabler, som även finns listade nedan. Användningen av areamedelvärden möjliggjorde för analys av storskaliga gradienter mellan land och hav, vilket var den drivande faktorn bakom sjöbris. Data lästes in från GRIB-filer för ERA5 och GFS och variablerna identifierades genom metadata (namn, enheter och rimliga väderintervall)

- 10 m vindkomponenter u och v
- 2 m temperatur T
- 2 m daggpunkt
- Lufttryck vid havsnivå
- Gränsskiktshöjd
- Total molnighet

Den nedladdade ERA5-data delades in i följande kategorier:

- Träningsmängd
- Valideringsmängd
- Testmängd

Träningsmängden omfattade insamlade data från 2010 till 2023 för sommarmånaderna juni, juli och augusti. Valideringsmängden omfattades av data från 2024 och testmängden av data från 2025 för samma månader som för träningsmängden. Den insamlade datan genomgick en förbehandlingsprocess där felaktiga och saknade datapunkter identifierades och hanterades. Saknade värden interpolerades där det ansågs lämpligt, medan datapunkter som avvek kraftigt från rimliga fysikaliska värden filtrerades bort. Indatavariablerna skalas från träningsdata baserat på medelvärde och standardavvikelse och transformationen appliceras sedan validerings och testdata. Därefter synkroniserades datamängderna tidsmässigt för att säkerställa en enhetlig tidsupplösning.

3.2 Konstruktion av egenskaper och målvariabler

För att möjliggöra effektiv modellering konstruerades ett antal egenskaper baserade på de ursprungliga meteorologiska variablerna, en så kallad variabelutveckling (*eng: feature engineering*). Variabelutveckling utgjorde en central del av metoden och användes för att omvandla rådata till informativa indatavariabler som bättre beskrev de fysikaliska processer som driver sjöbris.

3.2.1 Vindrepresentationer

För att skapa fysikaliskt tolkbara indata till modellen omvandlades vindkomponenterna till hastighet (V) och riktning (θ) enligt ekvation 24 och 25, där u och v var vindens komponent i öst-västlig respektive nord-sydlig riktning.

$$V = \sqrt{u^2 + v^2} \quad (24)$$

$$\theta = \text{atan2}(v, u) \quad (25)$$

Modellen använde kontinuerlig representation för riktning, enligt ekvation 26. Riktningen transformerades till kartesiska komponenter genom sin- och cos-komponenter vilket möjliggjorde för en kontinuerlig och stabil representation i modellen. Vidare möjliggjorde detta för användning av maskininlärning och eliminering av diskontinuiteter vid 0 grader och 360 grader.

$$\theta_{sin} = \sin(\theta), \theta_{cos} = \cos(\theta) \quad (26)$$

3.2.2 Land-Hav gradienter

För att kvantifiera de drivande mekanismerna bakom sjöbris beräknades gradienter mellan land och hav och användes sedan som egenskaper i modellen, genom differensen mellan temperatur (T), tryck (P) och vindhastighet (V) mellan land och hav. Ekvation 27 användes för att approximera den termiska drivningen, ekvation 28 för tryck och ekvation 29 för vindhastigheten. Dessa gradienter utgjorde centrala indata till modellen.

$$\Delta T = T_{land} - T_{hav} \quad (27)$$

$$\Delta p = p_{land} - p_{hav} \quad (28)$$

$$\Delta V = V_{land} - V_{hav} \quad (29)$$

3.2.3 Tidsrepresentation

Diskontinuiteter mellan den cykliska sjöbrisens och tiden eliminerades genom att tidsvariablerna t uttrycks i cyklisk form enligt ekvation 30. Tidsvariablerna inkluderades i cykliska egenskaper, vilket gjorde det möjligt för modellen att lära sig hantera dygnsvariationer i sjöbrisens utveckling.

$$t_{sin} = \sin\left(\frac{2\pi t}{24}\right), t_{cos} = \cos\left(\frac{2\pi t}{24}\right) \quad (30)$$

3.2.4 Dynamiska egenskaper

Tidsvariationer fångades genom beräkning av diskreta derivator enligt ekvation 31, för bland annat temperatur och vindhastighet. Dessa variabler användes sedan av modellen som indikationer på förändringstakten i atmosfären.

$$\Delta X_{1h} = X(t) - X(t - 1)$$

$$\Delta X_{3h} = X(t) - X(t - 3)$$

3.2.5 Sammansatta index

Sammansatta variabler för uppvärmningspotential och sjöbrispotential konstruerades och representerade icke-linjära samband. Dessa index användes som mått på sannolikheten för sjöbris. Ekvation 32 beskrev uppvärmningspotentialen (H) hos landmassan medan ekvation 33 gav ett mått på sjöbrispotential (SBF).

$$H = \Delta T(1 - C) \quad (32)$$

$$SBF = \frac{\Delta T}{V + \epsilon} \quad (33)$$

Där C var molnigheten och ϵ var en konstant som infördes för att undvika division med noll vid låga eller obefintliga vindhastigheter.

3.3 Regelbaserad klassificering av sjöbris

Eftersom observerade klassificeringar saknades, definierades målvariabler baserade på fysikaliska kriterier. Ett tidssteg klassificerades som sjöbris om följande villkor uppfylldes:

- Dagtid mellan 10 och 18
- Temperaturgradient $\Delta T > 1.5^\circ\text{C}$
- Bakgrundsvinden i *REGION_BOX* är mindre än 6 m/s
- Ökande gränsskiktsdjup över land relativt hav
- Förändring i vindriktningen över hav större än 12° eller ökning i vindhastighet över hav större än 0.7 m/s.

En svag sjöbrissignal definierades när dagtid, temperaturgradient och svag bakgrundsvind sammanföll med antingen en tydlig vindriktningsförändring eller en ökning av vindhastigheten. En stark sjöbrissignal krävde dessutom ökande gränsskiktsdjup samt både vindriktningsförändring och vindhastighetsökning.

Baserat på intensiteten definierades klassificeringsmålet som binärt enligt ekvation 34. 1 representerade förekomsten av sjöbris och 0 representerade ingen sjöbris.

$$y \in \{0, 1\} \quad (34)$$

3.4 Prognoshorisonter och målvariabler

Prediktion av framtida tillstånd möjliggjordes genom skapandet av målvariabler genom tidsförskjutning av ERA5-data enligt ekvation 35. Δt var den valda prognoshorisonten på 1 respektive 3 timmar.

$$Y(t) = X(t + \Delta t) \quad (35)$$

Följande mål definierades:

- Vindriktning (sin/cos)
- Vindförändring
- Sjöbrisklass ($t + \Delta t$)

För GFS-data användes dessa variabler direkt från prognosen utan tidsförskjutning.

3.5 LSTM-modell

I projektet användes en Recurrent Neural Network-modell (RNN) implementerad i PyTorch. Modellen konstruerades som en Long Short-Term Memory (LSTM) modell. Denna form av maskininlärning motiverades av att sjöbris är ett tidsberoende fenomen där nuvarande tillstånd påverkas av tidigare utveckling.

Modellen tränades på historisk ERA5-data, vilket bestod av en sekvens av tidigare observationer, som komplement till meteorologiska variabler vid en enskild tidpunkt. Modellen tränades för att förstå samband mellan bland annat tidigare temperaturgradienter, bakgrundsvind, molnighet och hur dessa tillsammans påverkade sannolikheten för sjöbris samt den framtida utvecklingen av vindhastighet och vindriktning. Samma nätverk tränades för två relaterade uppgifter:

- Binär klassificering av sjöbris (sjöbris/ingen sjöbris)
- Regression av vindförändringar vid 1 och 3 timmars prognoshorisont

3.5.1 Sekvensbaserad indatarepresentation

Indata organiserades som tidssekvenser. För varje prediktion användes ett fönster av tidigare timmar med längden 24. Sekvenslängden valdes för att fånga variationer på hela dygnet vilket var centralt för sjöbrisutveckling och innebar att modellen använde de senaste 24 tidsstegen som indata. Egenskaps-vektorn vid tidpunkt t betecknades X_t och indatasekvensen skrevs enligt ekvation 36:

$$X_t = (x_{t-L}, x_{t-L+1}, \dots, x_{t-1}) \quad (36)$$

Tillståndet vid nästa tidpunkt t skulle förutsägas av modellen. I datasetklassen konstruerades därför varje träningsprov som ett par bestående av en sekvens av tidigare observationer och motsvarande målvariabler för den efterföljande tidpunkten. Detta gjorde det möjligt för modellen att lära sig tidsberoenden, till exempel att en gradvis ökande temperaturskillnad mellan land och hav under förmiddagen ofta föregår en etablerad sjöbris senare under dagen.

3.5.2 Nätverksarkitektur

Modellen (SeaBreezeLSTM-V42) bestod av tre huvuddelar:

1. Ett LSTM-lager som behandlade tidssekvensen
2. Ett gemensamt fullt kopplat nätverk som omvandlade den temporala representationen
3. Två separata utgångsgrenar, ett för klassificering och ett för regression

LSTM-delen definierades med följande huvudparametrar:

- Antal lager: 2

- Dolda dimensioner (noder): 96
- Dropout: 0.25
- Batch first = True, vilket innebar att indata har formen (batchstorlek, sekvenslängd, antal egenskaper)

För en given sekvens producerade LSTM-lagret en dold representation för varje tidssteg. I modellen användes endast representationen från det sista tidssteget, då den antogs sammanfatta den information från sekvensen som var mest relevant för att förutsäga nästa tillstånd. Denna användes sedan som indata till nästa del av nätverket.

Representationen fördes sedan vidare till ett delat, fullt kopplat nätverk bestående av två linjära lager med aktiveringsfunktionen ReLU (Rectified Linear Unit), kompletterat med Layer Normalization och Dropout. Den delade representationen beskrevs enligt ekvation 37 där $g(*)$ representerade den gemensamma delen av nätverket.

$$\mathbf{z} = g(\mathbf{h}_L) \quad (37)$$

3.6 Modellens utdata

Från den gemensamma representationen $\mathbf{z}$ producerades två olika typer av utdata, en binär klassificering (sjöbris / ingen sjöbris) av sjöbris och en regression av framtida vindförändringar.

3.6.1 Binär klassificering av sjöbris

Den första utgångsgrenen bestod av ett linjärt lager som producerade ett enda logitvärde enligt ekvation 38. En logit definierades matematiskt som logaritmen av oddset för sannolikheten p att en viss händelse inträffar, detta enligt ekvation 38.

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (38)$$

Där p var sannolikheten för att händelsen inträffade. Den resulterande logiten beräknades med ekvation 39.

$$o_{cls} = \mathbf{w}_{cls}\mathbf{z} + b_{cls} \quad (39)$$

Där o_{cls} var den resulterande logiten, $\mathbf{z}$ var indatavariabler, b_{cls} var en förskjutnings-term (bias) som gjorde det möjligt för modellen att anpassa beslutsgränsen oberoende av indatavariablerna, och

0: ingen sjöbris

1: sjöbris

Även om den regelbaserade etiketteringen först skapade en intensitetsklass med värdena 0, 1 och 2 reducerades klassificeringsmålet till ett binärt problem vid träning. Klasserna 1 och 2 behandlades då som sjöbris, medan klass 0 behandlades som ingen sjöbris. Detta valdes för att möjliggöra för en robust operativ modell.

3.6.2 Regression av framtida vindförändringar

Den andra utgångsgrenen var ett linjärt lager som producerade sex kontinuerliga regressionsutgångar:

- Vindriktning efter 1 timme, representerade som $\sin(\theta_{1h})$ och $\cos(\theta_{1h})$.
- Förändring i vindhastighet efter 1 timme.
- Vindriktning efter 3 timmar, representerade som $\sin(\theta_{3h})$ och $\cos(\theta_{3h})$.
- Förändringar i vindhastighet efter 3 timmar.

Regressionsvektorn definierades enligt ekvation 40:

$$\widehat{y_{reg}} = [\widehat{\sin(\theta_{1h})}, \widehat{\cos(\theta_{1h})}, \widehat{\Delta V_{1h}}, \widehat{\sin(\theta_{3h})}, \widehat{\cos(\theta_{3h})}, \widehat{\Delta V_{3h}}] \quad (40)$$

Representation av cirkulär riktning med sinus och cosinus valdes för att eliminera diskontinuiteter. Utan denna omvandling framstod exempelvis 1 grad och 359 grader som långt ifrån varandra, när de egentligen var nästintill identiska.

3.7 Förlustfunktion

En sammansatt förlustfunktion valdes eftersom modellen tränades för både klassificering och regression samtidigt. Förlustfunktionen användes för optimering av modellen när den tränades.

3.7.1 Klassificeringsförlust

Klassificeringsdelen använde en binär kors-entropi med logiter, genom funktionen BCEWithLogitsLoss i koden. För ett enskilt träningsprov definierades förlusten enligt ekvation 41:

$$\mathcal{L}_{cls} = -[y \log(\sigma(o_{cls})) + (1 - y) \log(1 - \sigma(o_{cls}))] \quad (41)$$

Modellen använde sig av en klassviktning genom parametern pos_{weight} , eftersom antalet fall med och utan sjöbris inte var jämnt fördelade i träningsdata. Vikten beräknades som förhållandet mellan negativa och positiva exempel enligt ekvation 42. Detta medförde att modellen straffades hårdare när den missade positiva sjöbrisfall.

$$pos_{weight} = \frac{N_{neg}}{N_{pos}} \quad (42)$$

3.7.2 Regressionsförlust

Regressionsdelen använde medelkvadrattfel, MSE, för förlustberäkning, men med olika deltermer för riktning och hastighet. Förlusterna beräknades separat för prognoser för 1 och 3 timmar enligt ekvationerna 43 - 46. Riktningarna representerades av sina sinus- och cosinuskomponenter i ekvationerna.

$$\mathcal{L}_{dir,1h} = MSE(\hat{d}_{1h}, d_{1h}) \quad (43)$$

$$\mathcal{L}_{spd,1h} = MSE(\Delta\hat{V}_{1h}, \Delta V_{1h}) \quad (44)$$

$$\mathcal{L}_{dir,3h} = MSE(\hat{d}_{3h}, d_{3h}) \quad (45)$$

$$\mathcal{L}_{spd,3h} = MSE(\Delta\hat{V}_{3h}, \Delta V_{3h}) \quad (46)$$

I modellen vägdes sedan dessa samman enligt ekvation 47. Riktningen gavs något större vikt än hastighetsförändringen i regressionsdelen.

$$\mathcal{L}_{reg} = 1.5\mathcal{L}_{dir,1h} + 1.0\mathcal{L}_{spd,1h} + 1.5\mathcal{L}_{dir,3h} + 1.0\mathcal{L}_{spd,3h} \quad (47)$$

3.7.3 Total förlust

Den totala förlusten i träningen definierades enligt ekvation 48.

$$\mathcal{L}_{total} = 1.5\mathcal{L}_{cls} + 0.4\mathcal{L}_{reg} \quad (48)$$

Klassificeringen valdes att väga tyngre än regressionsuppgiften eftersom huvudmålet i den operativa tillämpningen var att avgöra om sjöbris förväntades uppstå eller inte.

3.8 Träning av modellen

Modellen tränades i PyTorch med optimeraren AdamW och använde en inlärningshastighet på 10^{-4} . Batchstorleken sattes till 32 och maximalt antal epoker till 35. För att stabilisera träningen användes gradient clipping med enligt ekvation 49, vilket begränsade gradienternas norm och minskade risken för instabilitet i LSTM-träningen.

$$\|g\| \leq 1.0 \quad (49)$$

För att stabilisera träningen och minska risken för att optimeringen fastnar i lokala minima användes en lärhastighetsreglering som reducerade lärhastigheten när valideringsförlusten slutade förbättras.

3.9 Kalibrering och beslutströskel

Modellen kalibrerade sannolikheterna efter träning genom Platt-skalning i koden. Platt-skalning innebar att de obehandlade logit- eller sannolikhetsvärdena från modellen omvandlades med en logistisk modell som tränades på valideringsdata enligt ekvation 50, där a och b bestämdes från valideringsmängden. Genom att använda detta skapades mjukare och mer stabila fördelningar. Detta implementerades Syftet var att modellens predikterade sannolikheter bättre skulle motsvara den faktiska sannolikheten för sjöbris.

$$p_{\text{cal}} = \frac{1}{1 + e^{-(az+b)}} \quad (50)$$

Vid operativ användning är användaren inte enbart intresserad av om modellen predikterar sjöbris eller inte, utan även av hur säker den var i sin prognos, vilket gjorde ovanstående extra viktigt.

3.9.1 Val av beslutströskel

Modellen använder en klassificeringsutgång, vilket är sannolikheten för sjöbris. En beslutströskel används för att omvandla sannolikheten till ett binärt beslut. Beslutströskeln valdes som den tröskel som gav högst precision under villkoret att recall, vilket är den andel verkliga sjöbrisfall, inte understeg 0.8. Sökningen gjordes med steg om 0.01 över intervallet 0.3 till 0.9. Denna metod användes eftersom modellen är tänkt att användas som beslutsstöd, där det är viktigare att inte missa för många sjöbrisfall än att maximera noggrannheten. För jämförelse beräknades även resultatet för flertalet fasta trösklar, vilket visade hur känslig modellen är för valet av beslutsgräns.

3.10 Utvärderingsmetod

När tröskeln valts utvärderades klassificeringsdelen med flertalet kompletterande mått.

3.10.1 Precision

Precision anger hur stor andel av de prognostiserade sjöbrisfallen som faktiskt är sanna. Precision beräknas enligt ekvation 51, där TP står för sanna positiva och FP för falska positiva.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (51)$$

3.10.2 Recall

Recall anger hur stor andel av de verkliga sjöbrisfallen som modellen korrekt hittade. Recall beräknas enligt ekvation 52, där FN är falska negativa.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (52)$$

3.10.3 F1-score

F1-score definierades som det harmoniska medelvärde av precision och recall. Detta beräknades genom ekvation 53 och är ett mått som är särskilt användbart när klasserna är obalanserade, vilket är fallet i detta projekt då sjöbris inte förekom vid varje tidssteg.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (53)$$

3.10.4 Balanserad träffsäkerhet (Balanced accuracy)

Eftersom klassfördelningen var skev användes även balanserad träffsäkerhet, vilket var ett mått som tog hänsyn till modellens träffsäkerhet i båda klasserna. Balanserad träffsäkerhet definierades som medelvärde av känsligheten och specificiteten enligt ekvation 54. Detta mått gav en mer rättvisande bild av klassificeringsprestandan än vanlig träffsäkerhet när positiva och negativa exempel inte var jämnt fördelade.

$$BA = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (54)$$

3.10.5 Confusion matrix

En confusion matrix användes för att tydligt visa klassificeringsresultatet. Antalet sanna positiva, sanna negativa, falska positiva och falska negativa sammanställdes i en matris. Detta är viktigt i operativa sammanhang eftersom olika typer av fel har olika konsekvenser. Ett falskt negativ innebär att modellen missar en sjöbris, medan ett falskt positiv kunde leda till att användaren förväntade sig en vindutveckling som inte skedde.

3.11 Modellval under träning

För att välja den bästa modellen användes inte enbart förlustvärdet, utan även operativt relevanta klassificeringsmått på valideringsdata. Efter varje epok beräknades bland annat följande:

- balanserad träffsäkerhet (BA)
- F1-score för sjöbris (F1)
- Precision
- Recall

Det sammanvägda mått som användes för modellval definierades i koden enligt ekvation 55.

$$S = 0.6 \cdot F1 + 0.4 \cdot BA \quad (55)$$

Detta mått visade att modellen i första hand optimeras för att ge en god träffsäkerhet på sjöbrisfallen, men att även balansen mellan klasserna togs i beaktning. För att undvika att modellen övertränades användes ett tidigt avslut om modellen inte förbättrade sitt valideringsresultat under sex efterföljande epoker. Den bästa tidigare modellen återställdes sedan efter stoppet.

3.12 Validering av regressionsutgångar

Utöver klassificeringen producerade modellen kontinuerliga prognoser för vindförändringar efter 1 och 3 timmar, vilket utvärderades separat.

3.12.1 Vindriktning

Modellens förutspådda sinus- och cosinuskomponenter för vindriktningen omvandlades tillbaka till en fysisk riktning i grader enligt ekvation 56. Därefter omvandlades riktningen till intervallet 0 till 360 grader. Detta gav en vindriktning som direkt kan tolkas vid operativ användning.

$$\hat{\theta} = \text{atan2}(\sin \hat{\theta}, \cos \hat{\theta}) \quad (56)$$

För att beräkna felet mellan verkliga och förutspådda riktningar användes ett cirkulärt felmått då vinklarna var periodiska. Vinkelfelet beräknades i ekvation 57 och definierades som den minsta absoluta skillnaden mellan två riktningar.

$$e_{\theta} = \left| \left((\hat{\theta} - \theta + 180) \bmod 360 \right) - 180 \right| \quad (57)$$

Det sammanfattade måttet som användes av modellen var cirkulärt medelabsolutfel (CMAE), vilket definierades enligt ekvation 58. Detta mått var mer rättvisande än vanlig MAE eftersom hänsyn togs till att endast 2 grader skiljde mellan 359 grader och 1 grad.

$$CMAE = \frac{1}{N} \sum_{i=1}^N e_{\theta,i} \quad (58)$$

3.13 Jämförelse mot persistensbaslinje

För att bedöma ifall modellen verkligen tillför värde jämfördes dess regressionsutgångar med en enkel persistensbaslinje, där persistens innebar att framtida vind antogs vara lika med det senaste observerade tillståndet. En sådan baslinje var enkel men oftast svår att slå i korttidsprognoser, speciellt när atmosfären förändrades långsamt.

För vindhastighet innebar detta att förändringen sattes till noll, vilket motsvarade antagandet att förändringen i vindhastighet sattes till noll enligt ekvation 59.

$$\Delta V_{\text{pers}} = 0 \quad (59)$$

För vindriktning innebar persistens att den framtida riktningen antogs vara lika med den aktuella riktningen enligt ekvation 60.

$$\theta_{\text{pers}}(t + \Delta t) = \theta(t) \quad (60)$$

Sedan jämfördes modellens fel med persistensens fel, både för 1 och 3 timmarshorisonten. Förbättringen definierades i modellen som skillnaden mellan baslinjens fel och modellens fel, enligt ekvation 61 för vindhastighet och ekvation 62 för vindriktning.

$$I_{\text{speed}} = MAE_{\text{pers}} - MAE_{\text{model}} \quad (61)$$

$$I_{\text{dir}} = CMAE_{\text{pers}} - CMAE_{\text{model}} \quad (62)$$

Ett positivt värde på ekvationerna 62 och 61 innebar att modellen presterade bättre än persistensbaslinjen, vilket var viktigt för att motivera användningen av en avancerad modell.

3.14 Fallstudie och operativ verifiering

Som komplement till den statistiska utvärderingen genomfördes även verifiering på enskilt utvalda dagar. Fallstudien genomfördes den 12-14 juli, vilket är relevant för den planerade framtida operativa användningen. För dessa dagar byggdes sedan timvisa sekvenser och modellen användes för att generera prognoser på samma sätt som vid operativ drift. För varje fallstudie jämfördes sedan följande:

- Verklig förekomst av sjöbris
- Modellens sannolikhet för sjöbris
- Binärt beslut efter vald tröskel
- Förutspådd förändring i vindriktning och vindhastighet
- Motsvarande värden från persistensbaslinjen

Genomförandet av fallstudier kompletterade de tidigare beräknade måtten med en mer praktisk bedömning av modellens beteende. Detta gjorde det möjligt att studera hur modellen reagerade på övergångar mellan morgonförhållanden och etablerad sjöbris, samt om den fångade relevanta förändringar i vindfältet under en tävlingsdag.

3.15 Sammanfattning av valideringsmetod

Valideringen är uppbyggd för att spegla modellens faktiska användning. Först tränades modellen på historiska data och därefter kalibrerades sannolikheterna på valideringsmängden och en operativ beslutströskel valdes. Slutligen utvärderades modellen på testdata

med mått som fångade både klassificering och regression, och resultaten jämfördes mot en enkel persistensbaslinje.

Denna valideringsstrategi gjorde det möjligt att bedöma av modellen ur två perspektiv samtidigt. Dels som en statistisk maskininlärningsmodell, dels som ett praktiskt verktyg för att stödja prognosbedömningar vid tävlingssegling.

4 Resultat

Resultaten från modellen som byggdes för att skapa en prognos för sjöbris vid kusten utanför San Pedro redovisades nedan. Dels värden som används för att utvärdera träningen av LSTM-modellen samt validering och verifiering. Modellens prognos för historiska dagar under 2025 redovisades också och jämfördes med vinddata från väderstationen 9410647 [19] vid samma tid och plats.

I detta avsnitt presenteras resultaten från den utvecklade LSMT-modellen. Först redovisas träningen av modellen med klassificeringsresultat, felmått för vindprognoser och kalibrering av sannolikheter. Därefter en ROC- och precisions-recall-analys och validering mot NDBC-data. Sedan resultatet av en fallstudie för en historisk specifik dag och en prognosvy för nästkommande dag.

4.1 Träning av modellen

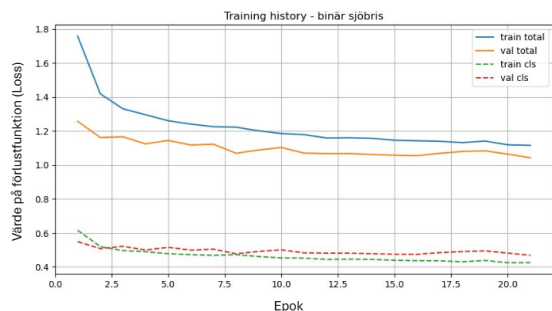
Under träningen användes en positiv viktning på 5.600. Detta resulterade i att ett sjöbrisfall värderades högre med en faktor 5.600 än ett icke-sjöbrisfall.

Automatiskt vald tröskel för sjöbrisfall från kalibrerad data:

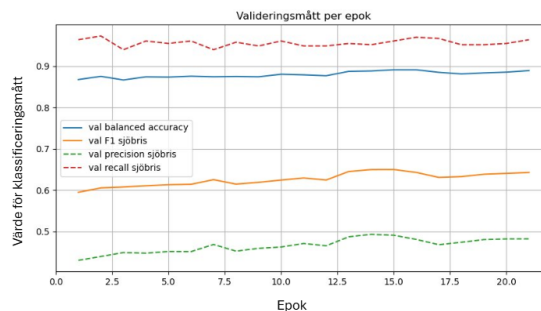
- Tröskel: 0.3
- Precision för sjöbris: 0.590
- Recall för sjöbris: 0.807
- F1 för sjöbris: 0.680
- balanserad träffsäkerhet: 0.853

Figur 5a visar hur modellens förlust förändrades under träningen. Den blåa linjen, vilket representerade träningsförlusten, minskade när modellen lärde sig olika mönster i datan. Den orangea linjen, vilket representerar valideringsförlust, visar hur modellen presterade på den data som den inte direkt tränades på. I grafen planar båda kurvorna ut, vilket betyder att modellen konvergerade. Om valideringsförlusten hade fått en kraftig ökning medan träningsförlusten hade konvergerat, skulle det vara ett tecken på att modellen övertränade sig själv.

Figur 5b visar balanserad träffsäkerhet, F1-score, recall, precision per epok. När kurvan med recall visar höga värden, innebär det att modellen hittade många fall av sjöbris. Precision visar falska sjöbrislarm och ett lägre värde betyder fler falska sjöbrisfall. Grafen visade en balans mellan hittade sjöbrisfall och falska positiva. När kurvorna stabiliseras sig tydde det på att träningen har blivit stabil.



(a) Tränings- och valideringsförlust



(b) Utveckling av balanserad träffsäkerhet, precision, recall och F1-score

Figur 5: Träningsförlopp för modellen

4.1.1 Klassificeringsresultat

Tabell 3 visar precision, recall och F1-score för sjöbris och ingen sjöbris. Modellen var bra på att identifiera dagar utan sjöbris med hög precision, 96 %. För sjöbris hade modellen en recall på 0.8066, vilket betydde att ungefär 81 % av sjöbrisfallen hittades av modellen. Precisionen för sjöbris var 0,5881, vilket betydde att ungefär 59 % av alla hittade sjöbrisfall var korrekta.

Klass	Precision	Recall	F1-score	Support
Ingen sjöbris	0.9629	0.8987	0.9297	1846
Sjöbris	0.5881	0.8066	0.6803	331
Accuracy			0.8847	2177
Macro avg	0.7755	0.8527	0.8050	2177
Weighted avg	0.9059	0.8847	0.8917	2177

Tabell 3: Klassificeringsrapport för verifieringsdata

Tabell 4 visar exakt hur många rätt och fel modellen hade när sjöbrisfall identifierades. 1659 fall klassades korrekt som ingen sjöbris, medan 64 av dem som klassades som ingen sjöbris var falska negativa och faktiska sjöbrisfall. 267 fall klassades som sjöbris, medan 187 av dem som klassades som sjöbris var falska positiva och var inga sjöbrisdagar. Det visar att det största felet var falska positiva sjöbrisfall.

	Predikterad ingen sjöbris	Predikterad sjöbris
Faktisk ingen sjöbris	1659	187
Faktisk sjöbris	64	267

Tabell 4: Förväxlingsmatris för träningsdata

Modellen uppnådde en balanserad träffsäkerhet på 0.85 vid verifiering mot ERA5-träningsdata för åren 2010-2023. Sjöbrisklassen erhöll precision 0.59, recall 0.81 och F1-värde 0.68. Detta visade att modellen identifierade en stor andel av sjöbristillfällena, men med ett relativt stort antal falska positiva klassificeringar.

4.1.2 Felmått för vindprognoser

Tabell 5 visar felet för prognosen. Vindprognoserna har medelabsolutfelet 0.32 m/s för 1 timmes prognos och 0.55 m/s för 3 timmars prognos. Vindriktningen hade ett cirkulärt medelabsolut fel på 12.7° respektive 16.1°. Resultaten indikerade att modellen hade en högre träffsäkerhet för kortare prognoshorisont.

Prognoshorisont	Circular MAE riktning	MAE vindhastighet
1 h	12.67	0.317
3 h	16.10	0.549

Tabell 5: Fel för prognostiserad vindriktning och vindhastighet

4.1.3 Kalibrering av sannolikheter

Tabell 6 visar hur modellens sannolikhet faktiskt motsvarade förekomsten av sjöbris. Mean prob är modellens genomsnittliga sannolikhet i ett intervall och event rate är hur ofta sjöbris faktiskt inträffade inom samma intervall. Låga sannolikheter var nära 0 i intervallet för event rate, medan högre sannolikheter hade ett högre värde inom intervallet.

Datadrivna nivågränser för sjöbris:

- Maximum för ingen sjöbrissignal: 0.0369
- Maximum för svag signal: 0.218
- Maximum för möjlig sjöbris: 0.3

Mean prob	Event rate	N	Min	Max	Usable
0.000 393	0.000 000	218	0.000 315	0.000 430	True
0.000 470	0.000 000	218	0.000 431	0.000 517	True
0.000 566	0.000 000	217	0.000 517	0.000 617	True
0.000 715	0.000 000	218	0.000 618	0.000 843	True
0.001 054	0.000 000	218	0.000 843	0.001 378	True
0.002 817	0.004 608	217	0.001 385	0.006 021	True
0.036 909	0.050 459	218	0.006 159	0.107 512	True
0.217 590	0.271 889	217	0.107 828	0.319 982	True
0.441 731	0.435 780	218	0.321 291	0.583 052	True
0.798 693	0.756 881	218	0.584 416	0.963 986	True

Tabell 6: Kalibreringsresultat för träningsdata

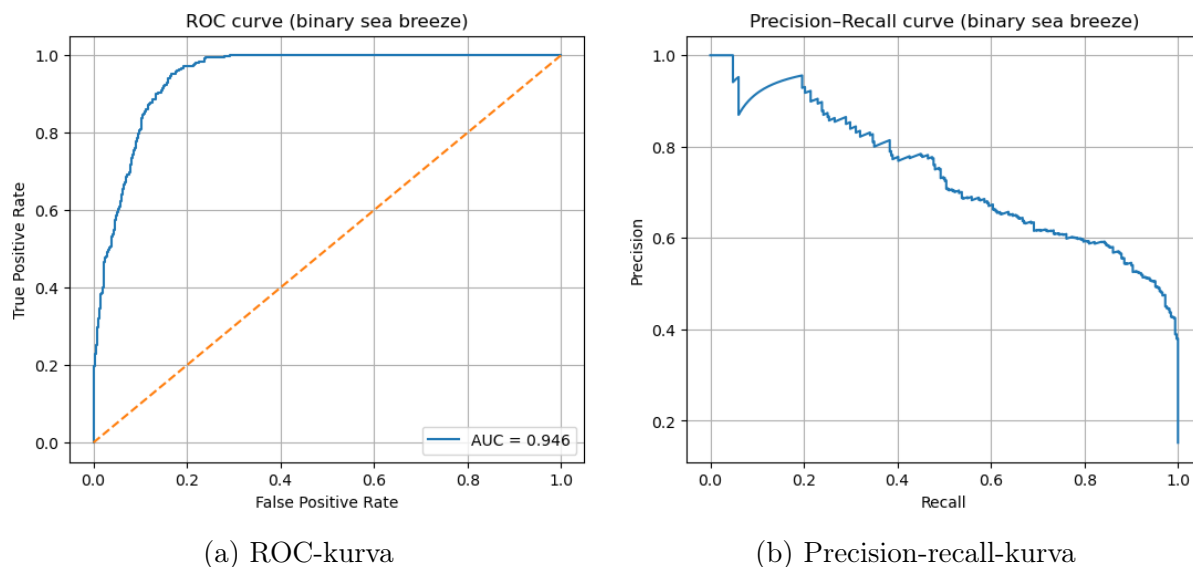
4.2 Validering och verifiering

I detta avsnitt testas modellens prognosförmåga genom en ROC-kurva och en precision-recall-kurva för att bestämma ett tröskelvärde för sjöbrissignalen. Därefter validerades modellens prognos mot NDBC vindobservationer.

4.2.1 ROC- och precision-recall-analys

Figur 6a visar hur bra modellen kan skilja på sjöbris och ingen sjöbris mellan olika trösklar. En kurva som sträcker sig mot det övre vänstra hörnet betyder att modellen har bra separationsförmåga. Figuren är inte beroende av något tröskelvärde, men den visar modellens generella klassifieringspotential.

Figur 6b visar sambandet mellan precision och recall över olika trösklar. Hög recall ger lägre precision och visar kompromissen mellan att hitta sjöbrisfall och att undvika falska positiva.



Figur 6: Modellens klassifieringsprestanda

Tabell 7 visar klassifieringsmått för de olika trösklarna i valideringsdatan. Trösklarna mellan 0.30 och 0.34 gav ett stabilt F1 värde, samtidigt som recall minskade och precisionen ökade. Inledningsvis antogs modellens tröskelvärde vara 0.3 men efter valideringen uppdaterades tröskelvärde till 0.32 efter optimeringen av recall och precision.

Tröskel	Precision	Recall	F1	Balanserad träffsäkerhet
0.30	0.5595	0.8519	0.6754	0.8659
0.31	0.5642	0.8369	0.6739	0.8605
0.32	0.5732	0.8278	0.6774	0.8586
0.33	0.5745	0.8157	0.6742	0.8537
0.34	0.5833	0.8036	0.6760	0.8504
0.87	0.9048	0.1148	0.2038	0.5563
0.88	0.8519	0.0695	0.1285	0.5337
0.89	0.8750	0.0634	0.1183	0.5309
0.90	0.8000	0.0363	0.0694	0.5173
0.91	0.6667	0.0181	0.0353	0.5083

Tabell 7: Klassifieringsmått för olika beslutströsklar på valideringsdata

4.2.2 Validering mot NDBC-data

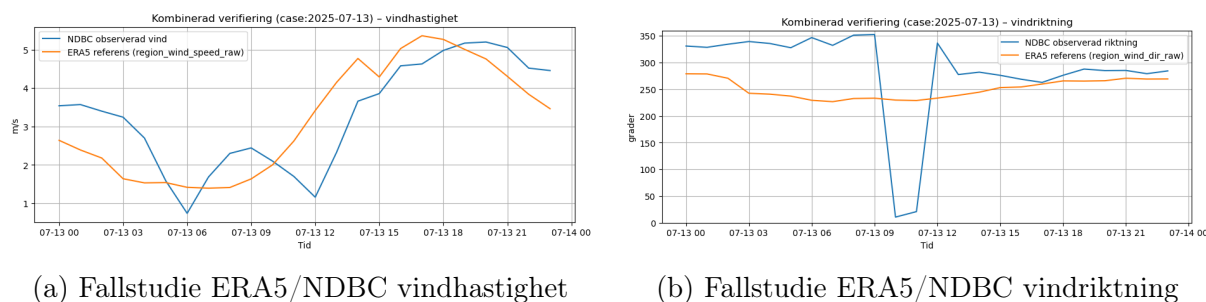
Tabell 8 jämför regional vind med observerad vind från en väderstation. Vindriktning, vindhastighet, observerad sjöbris och andra fysikaliska indikatorer visas i tabellen. Mellan kl. 14 och 15 gick sjöbrisobservationen från 0 till 1, vilket indikerar sjöbris, samtidigt som temperaturgradienten och sjöbrisfavorabiliteten ökade. Detta visar att de observerade data innehöll en tydlig sjöbrisssignal.

Tid	Vind (reg)	Dir (reg)	Vind (obs)	Dir (obs)	SB obs	ΔT	Fav.	Heating
00	2.64	279	3.54	331	0	-0.83	-0.96	-0.55
01	2.39	279	3.57	329	0	-1.07	-1.24	-0.55
02	2.18	271	3.40	334	0	-1.19	-1.24	-0.55
03	1.64	243	3.24	339	0	-1.02	-1.23	-0.55
04	1.53	241	2.70	336	0	-0.98	-1.16	-0.55
05	1.54	237	1.59	328	0	-0.88	-0.95	-0.54
06	1.42	229	0.74	347	0	-0.82	-0.80	-0.53
07	1.39	227	1.69	332	0	-0.70	-0.51	-0.48
08	1.41	233	2.30	351	0	-0.69	-0.50	-0.41
09	1.64	233	2.44	353	0	-0.65	-0.49	-0.36
10	2.00	230	2.10	11	0	-0.25	0.14	0.15
11	2.62	229	1.70	21	0	0.11	0.42	0.82
12	3.42	233	1.16	337	0	0.49	0.59	1.65
13	4.15	239	2.33	278	0	0.96	0.82	2.07
14	4.77	245	3.66	282	1	1.23	0.84	1.92
15	4.29	253	3.86	276	1	1.10	0.89	0.03
16	5.03	254	4.58	269	1	1.03	0.58	-0.43
17	5.36	260	4.63	263	1	0.91	0.38	-0.55
18	5.27	266	4.98	276	1	0.71	0.23	-0.55
19	5.01	265	5.17	288	1	0.23	-0.13	-0.55
20	4.76	266	5.20	285	1	-0.12	-0.40	-0.55
21	4.31	271	5.06	285	1	-0.31	-0.53	-0.55
22	3.84	269	4.52	279	1	-0.55	-0.72	-0.55
23	3.46	269	4.46	284	1	-0.83	-1.02	-0.55

Tabell 8: Observerade och regionala vindar samt indikatorer för sjöbrisutveckling

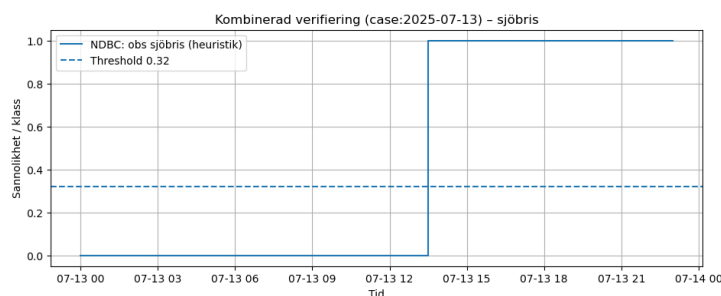
Figur 7a jämför vindhastigheten från ERA5 med faktiska observationer från NDBC och visade om ERA5 faktiskt representerade verkliga vindförhållanden. Båda datakällorna visade en vindökning under dagen.

Figur 7b jämför vindriktningen från ERA5 med faktiska observationer från NDBC och visade om ERA5 faktiskt representerade verkliga vindförhållanden. Båda datakällorna visade en vindvridning under sjöbrisintervallet vilket syns i båda datakällorna.



Figur 7: Jämförelse mellan ERA5 och NDBC

Figur 8 visar sjöbrissignal från observationsdata. Om signalen är 0 så finns det ingen sjöbriskaraktär och om signalen är 1 så finns det sjöbriskaraktär. Om sjöbrissignalen från vindstationsdatan sammanföll med sjöbrissignalen från ERA5 data så ströks kredibiliteten hos ERA5 datan.



Figur 8: Sjöbrissignal jämfört med NDBC-baserad klassificering

4.3 Fallstudie

I detta avsnitt presenteras resultaten från en fallstudie för ett specifikt dygn, där modellen testades genom att skapa en prognos och jämföra den med baslinjen.

4.3.1 Klassificeringsresultat för fallstudien

Tabell 9 visar modellens klassifikation under ett specifikt dygn. Om precision sjöbris var 1 innebar det att alla predikerade sjöbrisfall var korrekta. Recall sjöbris = 0.75 betydde att modellen hittade 3 av 4 sjöbrisfall i studien. F1 sjöbris = 0.857 visade balansen mellan precision och recall. Tabellen är bara för ett dygn och borde tolkas som en fallstudie och inget annat.

- balanserad träffsäkerhet: 0.875
- Precision sjöbris: 1
- Recall sjöbris: 0.75
- F1 sjöbris: 0.857

Klass	Precision	Recall	F1-score	Support
0	0.9524	1.0000	0.9756	20
1	1.0000	0.7500	0.8571	4
Accuracy			0.9583	24
Macro avg	0.9762	0.8750	0.9164	24
Weighted avg	0.9603	0.9583	0.9559	24

Tabell 9: Klassificeringsresultat

Tabell 10 visar hur exakt modellen klassade fallstudien som gjordes. 20 fall av sjöbris klassades korrekt och inga falska positiva fall. 3 fall av sjöbris klassades korrekt, med 1 fall av sjöbris som saknades. Modellen var konservativ i fallstudien jämfört med den större träningen av modellen.

	Predikterad ingen sjöbris	Predikterad Sjöbris
Faktisk ingen sjöbris	20	0
Faktisk sjöbris	1	3

Tabell 10: Förväxlingsmatris

4.3.2 Vindprognos i fallstudien

Tabell 11 visar en jämförelse mellan modellresultatet och baslinjen i fallstudien. Baslinjen som användes som referensmått för modellens prestanda utgjordes av en ihållande modell, där framtida tillstånd antogs vara lika med det aktuella tillståndet. Medelabsolutfelet för modellen var lägre än för baslinjen. För 1 timme var riktningsfelet 6.58 ° jämför med 11.11 ° för baslinjen och riktningsfelet 8.30 ° jämfört med 14.51 ° för baslinjen. Tabellen visade att modellen hade ett faktiskt prognosvärde.

Mått	Modell	Baseline
Circular MAE riktning (1h)	6.58	11.11
MAE hastighet (1h)	0.200	0.394
Circular MAE riktning (3h)	8.30	14.51
MAE hastighet (3h)	0.337	0.99

Tabell 11: Felmått för vindprognoser fallstudie

4.3.3 Tidsserieanalys av fallstudien

Tabell 12 visar facit för sjöbrissignalen, sannolikhet, observerad vind, prognostiserad vindhastighet och riktning timme för timme. Facit baserades på ERA5 data och prognostiseringen skedde på historiska GFS-data. De högsta sannolikheterna sammanföll med de tidsperioder då sjöbrissignalen enligt facit föreligger.

Tid	SB (facit)	Prob	Vind obs	Vind 1h	Dir obs	Dir 1h
00	0	0.001	2.64	2.31	279	272
01	0	0.001	2.39	2.03	279	273
02	0	0.001	2.18	1.84	271	272
03	0	0.001	1.64	1.33	243	270
04	0	0.001	1.53	1.38	241	256
05	0	0.001	1.54	1.49	237	254
06	0	0.001	1.42	1.48	229	258
07	0	0.002	1.39	1.61	227	256
08	0	0.007	1.41	1.70	233	258
09	0	0.054	1.64	2.02	233	258
10	0	0.182	2.00	2.47	230	255
11	1	0.543	2.62	3.35	229	247
12	1	0.596	3.42	4.06	233	247
13	1	0.376	4.15	4.50	239	252
14	0	0.236	4.77	4.97	245	255
15	1	0.094	4.29	4.42	253	259
16	0	0.021	5.03	5.16	254	262
17	0	0.004	5.36	5.25	260	266
18	0	0.002	5.27	4.96	266	271
19	0	0.001	5.01	4.61	265	273
20	0	0.001	4.76	4.30	266	273
21	0	0.001	4.31	3.84	271	275
22	0	0.001	3.84	3.41	269	275
23	0	0.001	3.46	3.03	269	274

Tabell 12: Jämförelse mellan observerad och prognostiserad vind samt sjöbrissignal

I tabell 13 visas skillnaden mellan den observerade och den prognostiserade vinden. Ett negativt värde betydde att modellens vindhastighet eller vindriktning var lägre än den observerade vid samma tidpunkt.

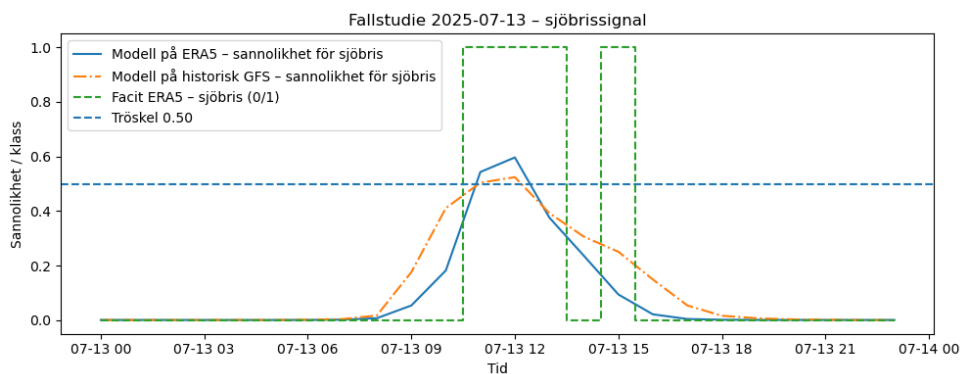
Tid	Vind fel	Dir fel
07	0.22	29
08	0.29	25
09	0.38	25
10	0.47	25
11	0.73	18
12	0.64	14
13	0.35	13
14	0.20	10
15	0.13	6
16	0.07	8
17	-0.11	6
18	-0.31	5
19	-0.40	8

Tabell 13: skillnaden mellan den observerade vinden och den prognostiserade vinden

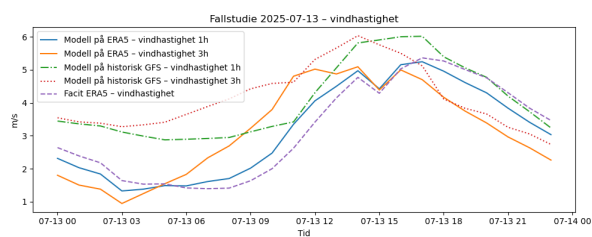
Figur 9a jämför modellens sannolikhet med ERA5-facit och GFS-baserad prognos. Grafen visade när modellen trodde att sjöbris var som mest sannolik och visade om modellen kunde plocka upp rätt tid inom de streckade gröna sjöbrissignalerna för ERA5- facilit.

Figur 9b visar vindhastigheten över ett dygn för både observerade och modellerade data. Den streckade lila linjen representerar ERA5 och användes som facit. De heldragna linjerna visar modellens prognoser baserade på ERA5 som indata, där blå linje motsvarade 1 timmesprognos och orange linje motsvarar 3 timmesprognos. De streckade gröna och röda linjerna motsvarande prognoser baserade på historiska GFS-data för 1 timme respektive 3 timmar. Grafen visar om modellen förbättrade och justerade GFS till en mer exakt vindhastighet. Ifall modellkurvan låg närmare ERA5 än GFS betydde det att modellen förbättrade och justerade GFS.

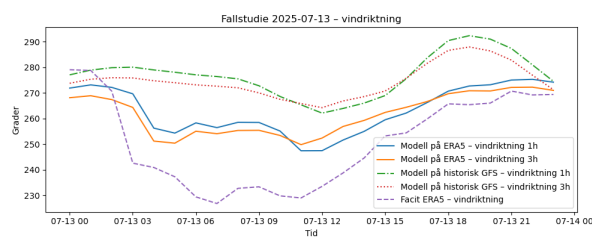
Figur 9c jämför vindhastigheten mellan ERA5, GFS och modellprognosen. Samma sak gällde för modellen, som förbättrade och justerade GFS till en mer exakt vindriktning. Om modellkurvan ligger närmare ERA5 än GFS betyder det att modellen förbättrade och justerade GFS.



(a) Fallstudie ERA5/GFS sjöbrissignal



(b) Fallstudie ERA5/GFS vindhastighet



(c) Fallstudie ERA5/GFS vindriktning

Figur 9: Jämförelse mellan ERA5, GFS och modell

4.4 Prognosvy för nästkommande dag

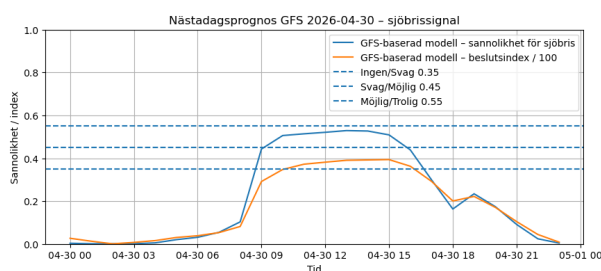
Tabell 14 visar den operativa prognosen timme för timme och innehåller sannolikhet, index, signalnivå, vindhastighet och vindriktning. Sannolikheten för sjöbris ökade från morgonen till mitt på dagen.

Tid	Prob	Index	Nivå	Vind 1h	Dir 1h	Vind 3h	Dir 3h	Vind op	Dir op
00	0.003	2.6	ingen	3.55	240	3.62	244	3.59	242
01	0.001	1.3	ingen	3.70	235	3.72	239	3.71	237
02	0.001	0.1	ingen	3.94	231	3.95	235	3.95	233
03	0.001	0.7	ingen	3.16	226	3.18	230	3.17	228
04	0.005	1.6	ingen	2.77	216	3.02	226	2.92	221
05	0.020	3.0	ingen	2.54	211	2.93	225	2.77	218
06	0.031	3.8	ingen	2.71	208	3.17	223	2.99	216
07	0.054	5.3	ingen	2.90	208	3.30	224	3.14	216
08	0.103	8.2	ingen	3.17	206	3.56	224	3.41	215
09	0.443	29.1	svag	3.31	206	3.71	225	3.55	216
10	0.506	34.8	möjlig	3.49	209	3.86	227	3.71	218
11	0.514	37.2	möjlig	3.65	211	3.92	229	3.81	220
12	0.521	38.2	möjlig	2.86	211	3.08	229	2.99	220
13	0.529	39.0	möjlig	2.37	214	2.68	231	2.55	222
14	0.527	39.2	möjlig	2.17	215	2.51	232	2.38	223
15	0.509	39.4	möjlig	2.90	220	3.25	236	3.11	228
16	0.439	36.3	svag	3.98	228	4.23	241	4.13	234
17	0.299	29.2	ingen	5.14	237	5.28	246	5.22	242
18	0.163	20.1	ingen	4.80	244	4.83	248	4.82	246
19	0.234	22.2	ingen	4.71	244	4.84	248	4.79	246
20	0.174	17.1	ingen	4.59	243	4.66	248	4.63	246
21	0.091	10.4	ingen	3.97	242	3.92	246	3.94	244
22	0.024	4.4	ingen	3.77	235	3.81	242	3.79	239
23	0.004	0.8	ingen	3.83	227	3.88	235	3.86	231

Tabell 14: Timvis prognos av sjöbris och vind i lokal tid(avrundade värden)

4.4.1 Sjöbrissignal

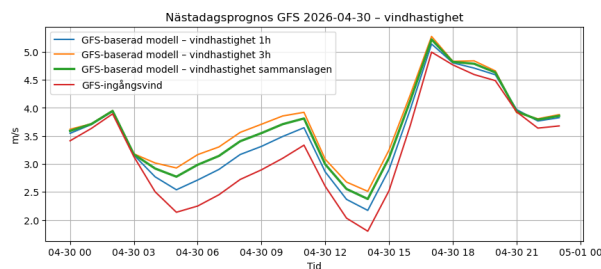
Figur 10 visar sannolikheten för sjöbris nästkommande dag. Grafen används som en visuell tolkning och visade tidsfönster med störst sannolikhet för sjöbris.



Figur 10: Timvis prognos av sjöbrissannolikhet för nästkommande dag

4.4.2 Vindhastighet

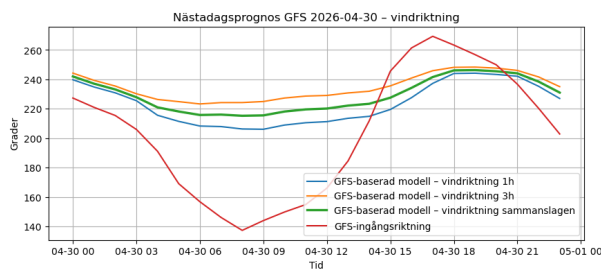
Figur 11 visar prognostiserad vindhastighet för 1 och 3 timmar framåt. Grafen hjälpte till att tolka om tidsfönster för sjöbrischans medförde en ökning av vindhastigheten. Vindhastigheten i tidsfönstret sjunker, vilket skulle indikera att det inte kommer att inträffa någon sjöbris under denna dagen.



Figur 11: Prognostiserad vindhastighet för 1 och 3 timmars prognoshorisont

4.4.3 Vindriktning

Figur 12 visar prognostiserad vindriktning för 1 och 3 timmar framåt. Grafen hjälpte till att tolka om tidsfönster för sjöbrischans medförde en vindvridning som gick åt höger. Vindriktningen i tidsfönstret gick lite åt vänster, vilket indikerade att det inte kom att inträffa någon sjöbris under denna dag.



Figur 12: Prognostiserad vindriktning för 1 och 3 timmars prognoshorisont

4.5 Sammanfattning av resultat

Sammanfattningsvis visade resultatet att modellen identifierade många sjöbrisfall med en recall på 0.81 för sjöbrislassen och att precisionen på 0.59 gav en del falska positiva. Vindprognosen för en timmes prognos framåt var mest träffsäker med MEA 0.317 m/s för vindhastigheten och circular MEA 12.67 ° för vindriktningen. Modellprognosen gav lägre fel än baslinjen.

5 Diskussion och analys

För att kunna utvärdera om den modell som utvecklats kan användas för lokala prognoser i San Pedro analyseras resultaten som presenteras i avsnitt 5. Arbetets syfte var dels att skapa en ökad förståelse för den lokala sjöbrisens och dels att utveckla en datadriven modell som kan identifiera och prognostisera sjöbris utifrån relevanta meteorologiska variabler.

Genom att kombinera modellens resultat med den litteraturstudie som genomförts av de fysikaliska processer som styr sjöbrisens uppkomst och utveckling, skapas goda förutsättningar för att tolka resultaten. Detta möjliggör analys av både modellens prestanda och förmåga att representera de underliggande fysikaliska processerna. I denna del analyseras därför resultaten utifrån modellens noggrannhet och begränsningar, modellens koppling till fysikaliska mekanismer och hur användbar prognosen som modellen skapar är i sammanhanget.

5.1 Modellens prestanda

LSTM-modellens prestanda utvärderas genom analys av träningsresultaten, genom att göra en validering mot observationsdata från NDBC. Utöver detta görs en fallstudie för att undersöka i detalj modellens förmåga att återge enskilda sjöbrisdagar. Detta för att bedöma hur väl modellen klarar av att identifiera och prognostisera sjöbrisens och då beskriva de lokala vindförhållandena.

5.1.1 Motivering av modellval

Användningen av en LSTM-modell motiveras genom att sjöbris är ett tidsberoende fenomen. Modellens uppgift är att förstå hur sjöbrisens utveckling över tid och inte enbart analysera ett enskilt meteorologiskt tillstånd. En traditionell statisk modell som enbart använder en tidsrad i taget riskerar att missa viktiga mönster som föregår utvecklingen av sjöbris. Modellens gemensamma prediktionsstruktur motiveras av att upptäckten av sjöbris och vindprognos är två delar av samma fysikaliska process. Om modellen lär sig identifiera mönster som leder till sjöbris bör dessa även vara relevanta för att förutsäga förändringar i vindriktning och vindhastighet.

5.1.2 Utvärdering av träningen på modellen

Eftersom modellen tränas på ERA5-data, som är en reanalys där observationer assimileras, kan ERA5 ge en mer representativ bild av de atmosfäriska förhållandena än GFS-prognoser. Detta gör att modellen i träningen exponeras för ett mer korrekt samband mellan de meteorologiska variablerna från sjöbrisens uppkomst och utveckling. Detta är en viktig förutsättning för modellens prestanda och gör den bättre på att förutsäga relevanta mönster kopplade till sjöbris i atmosfären.

Klassificeringsprestandan, det vill säga hur väl modellen identifierar om det är sjöbris eller inte, kan analyseras genom värdena på precision, recall och F1-score som presenteras i tabell 3. För klassen ingen sjöbris fås en hög precision på 0,97, vilket innebär att modellen i de flesta fall identifierar frånvaro av sjöbris korrekt. Medan för sjöbrisklassen är precisionen lägre, 0.59, vilket innebär att ungefär 40% av de fall som modellen klassar som sjöbris egentligen inte är det, det vill säga falska positiva som var 188 stycken. Detta syns

även i förväxlingsmatrisen i tabell 4 där antalet falska positiva är relativt högt. Modellen klassificerar alltså att det är sjöbris oftare än vad som faktiskt observeras, vilket innebär att modellen är känslig för sjöbrisfall. Samtidigt visar recallvärdet på 0,82 att modellen lyckas med att identifiera en stor andel av tiderna med sjöbris, vilket gör att få fall missas. Detta syns i antalet falska negativa som endast var 59.

Obalansen mellan precision och recall är kopplad till den positiva viktningen som appliceras på klassen sjöbris under träningen. Detta gör att modellen vill minimera antalet missade fall av sjöbris snarare än att minimera falska positiva. Ur ett meteorologiskt perspektiv är detta ett önskvärt beteende, då det ofta är mer kritiskt att missa en sjöbris än att överskatta hur ofta den utvecklas. På grund av att projektet är inriktat på segling och då mer kustnära prognoser kan ett uteblivet sjöbrisfall skapa större konsekvenser än falska positiva prognoser. Resultaten indikerar också att sjöbris är ett vanligt förekommande fenomen i det område som observeras. På grund av detta motiveras modellens känslighet.

Tabell 5 visar felen för den prognostiserade vindriktningen och vindhastigheten, som är låga i sammanhanget och visar att modellen ger ett resultat med hög noggrannhet. Resultaten tyder på att modellen identifierar samband som är förenliga med de fysikaliska processer som styr sjöbrisen. Den låga felnivån i vindriktning indikerar att modellen återger sjöbriscirkulationen som drivs av temperaturskillnader mellan land och hav. Modellen ger en mer exakt prognos för den kortare prognoshorisonten på 1 timme än för den längre på 3 timmar. Den högre träffsäkerheten för kortare prognoshorisonter är förväntad, då sjöbrisen är en mesoskalig process vars utveckling sker över timmar och påverkas mycket av lokala förhållanden.

Kalibreringsanalysen i tabell 6 visar att modellens sannolikheter i stort motsvarar den observerade frekvensen av sjöbrisfall, särskilt vid högre sannolikheter. Den relativt låga tröskeln på 0,32 gör att modellen är känslig för svagare signaler, vilket också bidrar till det höga antalet falska positiva. Modellens höga igenkännlighet för sjöbris är mycket relevant då detta även innebär att den blir känsligare för de små temperaturdifferenser som driver sjöbrisen.

5.1.3 Analys av Fallstudie

I resultaten redovisas en fallstudie som möjliggör en mer ingående analys av en specifik dag, 13 juli, där modellens prognoser jämförs med observerad vind från NDBC-data. I tabell 12 presenteras vindhastighet och vindriktning med en prognoshorisont på 1 timme tillsammans med motsvarande observationer. Klassificeringsresultaten och felen som identifieras i fallstudien är tillräckliga för att kunna ha förtroende för fallstudien.

I tabellen identifierar modellen sjöbris under tidsperioden 11–13 och har en sannolikhet kring 50% att uppstå. Detta stämmer överens med litteraturen, där sjöbrisen förväntas utvecklas under förmiddagen i takt med att temperaturgradienten som driver processen ökar. Under den aktuella dagen är vinden svag på morgonen, vilket är förhållanden som gynnar en ren sjöbris. En sådan utvecklas gradvis under dagen och leder till en successiv ökning av vindhastigheten i sjöbrisflödet. Detta framgår av tabellen, både i den observerade vindhastigheten och i LSTM-modellens vindhastighet. Klassificeringsresultaten som visas i tabell 9 indikerar en hög precision för klassen sjöbris men en lägre recall. Detta betyder att modellen, i detta fall, främst identifierar tydliga sjöbrissignaler, men kan missa

svagare fall. Detta kan bero på att sjöbriscirkulationen kvarstår senare under dagen även om modellen inte klassificerar det som sjöbris. I tabell 12 syns det att vinden fortsätter att bygga i styrka efter klockan 14, men klassificeras inte som sjöbris. Vilket tyder på att modellen har svårare att identifiera svagare fall eller de mer utvecklade faserna av cirkulationen, där signalen blir mindre tydlig. Samtidigt är det under denna period som modellen minimerar skillnaden mellan observation och prognos, detta avläses i tabell 13.

Vidare visar jämförelsen att modellen i detta fall ger en mer högervriden vindriktning jämfört med observationerna. En möjlig anledning till detta är skillnader i mätförhållanden. Om NDBC-datan representerar vind nära havsytan, där friktion påverkar riktningen mer, medan modellen visar ett mer idealiserat flöde. Vidare kan det tyda på ett systematiskt fel i hur modellen tar hänsyn till friktion. Observationerna visar även en gradvis vridning mot höger under dagen i takt med att sjöbrisen ökar i styrka. Detta är förenligt med den påverkan som corioliseffekten har på sjöbrisen. Modellen visar också denna trend, men i en lägre utsträckning, vilket kan vara ett tecken på att modellen inte fullt ut fångar utvecklingen av cirkulationen över tid.

Tabell 11 jämför MAE-felen för LSTM-modellen och en baslinje för både 1 och 3 timmars tidshorisont. Det kan tydligt avläsas att absolutfelet är mindre för modellen än för baslinjen för både vindhastighet och vindhastighet. Förbättringen är större för den längre tidshorisonten på 3 timmar, då felet går från 0.99 m/s till 0.34 m/s. Eftersom sjöbrisen är en dynamisk process visar detta på att modellen har förmåga att representera vindförhållanden bättre än baslinjen.

Sammanfattningsvis visar fallstudien att modellen identifierar tiden för etableringen av sjöbrisen och hur den utvecklas i styrka, men att avvikelser i riktning förekommer. Den visar också att modellen tenderar att missa svagare eller faser senare i utvecklingen av sjöbriscirkulationen. Jämförelserna med baslinjen visar dock på en tydlig förbättring, vilket gör modellen relevant för prognoser i området.

5.2 Fysikalisk tolkning av resultaten

Modellen är statistisk och tränas på historiska data med sjöbris. Detta innebär att modellen inte kan fysiken bakom processen utan lär sig mönster i temperatur och vind, vilket i sin tur kan påverka resultaten, då bland annat valet av geografiska områden och meteorologiska variabler har stor påverkan. Under detta avsnitt diskuteras hur dessa val samt modellens oförmåga att förstå fysikaliska processer kan ha skapat begränsningar i prognosen.

5.2.1 Geografiska områden

Studieområdet delas in i tre geografiska områden: en större region som täcker både land och hav, samt ett mindre landområde och ett mindre havsområde. Både storleken och positionen på dessa områden påverkar modellens resultat. De variabler som används i modellen som temperaturskillnad, tryckskillnad och vindskillnad, definieras som skillnaden mellan landområdet och havsområdet. Valet av storlek på boxarna påverkar också resultaten. Stora boxar ger inte utslag för relevanta signaler och missar då temperatur och tryckgradients som påverkar cirkulationen, medan små boxar ger brus som modellen har svårt att hantera.

Kustlinjen är dock mer komplex än vad som representeras av land- och havsboxarna. Detta gör att kustens exakta utformning och lokala topografi inte tas i beräkning. Modellen lär sig hur dessa faktorer påverkar vinden i respektive box, men kan inte explicit representera deras inverkan på sjöbriscirkulationens utveckling. I områden med komplex kustgeometri, som i San Pedro där kustlinjen består av bukter och höjdskillnader, kan sjöbrisen initialt vara mycket lokal. På grund av att dessa lokala variationer inte detekteras fullt ut i de valda områdena kan modellen ha svårt att korrekt prognostisera den tidiga utvecklingen av cirkulationen. Detta visualiseras i tabell 13 där skillnaden mellan modellens prognos och observationer är som störst i initieringen av sjöbrisen.

Fortsättningsvis så baseras modellens representation av sjöbrisen på temperaturskillnaden mellan land- och havsboxen. Under sjöbrisens utveckling, när sjöbrisfronten successivt rör sig längre in över land kan den drivande temperaturskillnaden förskjutas från kustlinjen till områden längre in över land. I San Pedro finns det även en bergskedja en bit in över land, detta leder till ytterligare uppvärmning av landmassan och förstärker uppflödet vid sjöbrisfronten. Vilket i sin tur förstärker sjöbriscirkulationen. denna uppdelning kan därför vara en bidragande orsak till att modellen underskattar sjöbrisens utbredning och styrka, särskilt senare under dagen.

Analysen av valet från de geografiska områdena kan sammanfattas med att de valda regionerna i modellen främst är anpassade för att prognostisera sjöbrisen under mitten av dagen. Initieringen av sjöbrisen sker med små lokala temperaturskillnader som i modellen kan framstå som brus. När sjöbrisen är etablerad sker drivningen på en större skala som stämmer överens bättre med de valda områdena. Modellen ger då en mer representativ bild av sjöbrisens styrka. Under senare delar av dagen, när cirkulationen växer, så finns det en risk att modellen underskattar sjöbrisens styrka. Detta syns i fallstudien, tabell 13, visar modellen förmåga att överskatta vindhastigheten på förmiddagen, men underskatta den på eftermiddagen. Detta stärker det tidigare resonemanget om att modellens geografiska uppdelning påverkar dess förmåga att förutse de olika faserna av sjöbrisens utveckling.

5.2.2 Övriga kopplingar till teorin

Sjöbrisen är en komplex, flerdimensionell process vilket gör den svår att modellera i detalj och innebär att prognoser som är baserade på förenklade metoder kan bli missvisande. På grund av att LSTMA-modellen bygger på statistik samband kan den återge mönster i sjöbriscirkulationen utan att lära sig de fysikaliska processerna. Detta gör att modellens prestanda beror på hur väl dessa processer visas i de variabler som används. Till exempel bakgrundsvinden har en stor påverkan på sjöbriscirkulationen. Vid en svag synoptisk vind etableras en ren sjöbris, detta kan modellen identifiera väl. Detta kan ses i fallstudien där låg bakgrundsvind gav ett relativt litet fel mellan modell och observationer. Vid starkare bakgrundsvindar kan dock sjöbriscirkulationen förstärkas och deformeras, det vill säga sjöbris kan uppstå trots stark bakgrundsvind. Detta leder till mer komplexa flödesmönster och modellen har då svårare att klassificera sjöbris, då signalen kanske avviker från de mönster som modellen tränats på.

Den vertikala strukturen och turbulensen i gränsskiktet är en annan utmaning. Sjöbrisen är en tredimensionell cirkulation där uppvärmning av markytan driver en vertikal transport. Detta skapar turbulens och är en oförutsägbar process som påverkar sjöbrisen. Modellen baseras främst på meteorologiska variabler nära markytan, såsom temperatur,

vind och lufttryck, medan information om atmosfärens vertikala struktur är begränsad. Detta kan leda till att modellen missar vissa sjöbrisfall eller ger avvikelser i vindriktning, särskilt när cirkulationen utvecklas över tid.

Sammanfattningsvis registrerar modellen främst de tydliga och etablerade sjöbrisfallen, medan mer komplexa fall, där till exempel bakgrundsvind och vertikal struktur inte följer de vanliga mönstren, är svårare att förutsäga.

5.3 Begränsningar och Osäkerhet

Resultatet i studien påverkas av osäkerheter inom flera områden. I detta avsnitt diskuteras hur de kopplas till dataupplösning, modellens uppbyggnad och flera metodval.

5.3.1 Dataupplösning

ERA5 som modellen har använt för träning är en realanalys, vilket gör den mer realistisk än prognosdata, men den är heller inte helt exakt och innehåller modellfel inom sin egen ram. Upplösningen över rutnätet som analyseras är 640 km^2 och därför kan ERA5 ha svårigheter att registrera småskaliga fenomen. Eftersom sjöbris är ett mesometeorologiskt lokalt fenomen som varierar beroende på lokala temperaturgradienter så har ERA5 svårt att beskriva exakta lokala meteorologiska variationer runt San Pedros kust.

Modellen använder GFS-prognoser som indata när den operativa prognosen ska byggas. Prognoserna från GFS är också tillämpade på en större skala och kan missa lokala ändringar i atmosfären. Alla prognoser kan innehålla fel just för att de är en framtida prognos, medan ERA5 data är korrigerade i efterhand. Därför kan det vara problematiskt att modellen tränar på data från en realanalys, men data som används i modellen är från en prognos. Olika slags data behöver inte vara något problem då modellens fel i GFS-prognosdata kan korrigeras, men felet kan också förstärkas, vilket applicerar en viss osäkerhet i modellens resultat.

NDBC-datan som används som en referens för att kunna utvärdera modellen kan också vara en felkälla. Den viktiga slutsatsen från detta är att ERA5-data kan representera de verkliga förhållandena på området. Dock syns en tydlig skillnad mellan ERA5 och NDBC i figur 7b där den observerade vindriktningen minskar hastigt klockan 9–12. Vindobservationerna tas från en station nära hamnen i San Pedro och därför kan vinddatan från stationen innehålla brus, eftersom topografin runt stationen kan påverka vinden kraftigt, speciellt om stationer sitter i lä om berg, högre byggnader eller liknande. Observationerna sker också punktbaserat inom områdesboxarna och detta tillför även ett lager av osäkerhet då modellen jämför data med olika upplösningar. Detta innebär inte nödvändigtvis att data från källan är oförlitliga, men det kräver medvetenhet vid användning.

5.3.2 Modellens begränsning

LSTM-modellen lär sig statistiska samband snarare än fysiska samband. Därför fungerar modellen som bäst när indatan liknar träningsdata till så stor del som möjligt. Vidare kan det bli svårt för en LSTM-modell att lära sig svåra sjöbrissignaler som är mer ovanliga. Det kan ses i resultatet att modellen har vissa svårigheter att identifiera dagar med sjöbrisindikationer som inte är lika tydliga och medför osäkerhet i klassificeringen av sjöbrisfall.

För att definiera sjöbris används tröskelvärden, vilket blir en förenkling av en verklig process. Ibland är övergången mellan sjöbris och icke-sjöbris ofta långsam och gradvis över en hel dag. Därför kan icke-sjöbrisdagar klassificeras felaktigt. Modellen använder sig av ett visst antal valda features. De utvalda meteorologiska parametrarna kan inte själva beskriva sjöbrisens utveckling. Parametrar som rör toppografi och atmosfärisk vertikal transport representeras inte i modellen och riskerar att träningen av modellen blir för simpel.

5.3.3 Klimatförändringens inverkan på framtidens prognoser

LSTM-modellen är tränad på historisk data från perioden 2010-2023, vilket innebär att dess prognosförmåga bygger på att de samband som identifieras i datan kvarstår över tid. Eftersom sjöbrisen är en temperaturdriven process finns det en potentiell påverkan av klimatförändringar. Den globala uppvärmningen kan påverka sjöbrisens cirkulation i styrka och frekvens. Detta gör att modellens relevans kan minska över tid, men denna effekt är svår att mäta eller visa för detta arbete.

5.4 Operativ användbarhet

Resultaten visar att modellen kan producera en potentiell operativ sjöbris för San Pedros kust. Den operativa prognosen är en kombination av GFS-prognosen och den tränade modellen som har lärt sig att uppskatta vindförhållanden och sjöbrisförhållanden på en lokal nivå.

Jämförelserna mot fallstudien visar att modellen ger ett lägre fel än baslinjen när det kommer till vindriktning och vindhastighet. Detta tyder på att modellen tillför information som kan komplettera en GFS-prognos för att göra den mer realistisk. Förbättringarna som kan tydas är störst för den prognosen med kortare tidsuppskattning. Det är förväntat till en viss del eftersom sjöbrisen är en mesoskalig process som sker under relativt kort tid under dygnet.

Med identifieringen av sjöbrisfall ger modellen hög recall, vilket innebär att många fall av sjöbris kan identifieras, men modellen har också lägre precision, vilket innebär en del falska positiva fall. Vilket i sin tur innebär att många fall av sjöbris identifieras, men några av de identifierade fallen är inte riktiga sjöbrisdagar. Däremot minskar detta risken att faktiska sjöbrisdagar missas av modellen. För en operativ användning kan detta vara en fördel då seglingen i San Pedro under det valda tidsintervallet har en hög förekomst av sjöbris. En prognos kan inte exakt förutsäga framtida förhållanden och om modellen antyder en sjöbrisdag kan seglarna observera tidiga tecken tidigt på dagen som skulle indikera motsatsen.

Modellen kan även beskriva etableringen av sjöbrisen under dagen från den timvisa prognosen. Däremot finns det en viss osäkerhet i dagar med svagare sjöbrissignaler eller senare faser av sjöbriscirkulation. Just därför ska prognosen tolkas med en viss försiktighet och inte som en exakt prognos.

Om man ser på det mest användbara från modellen är det framför allt sjöbrissannolikheten, vindhastigheten och vindriktningen. Sjöbrissannolikheten är ett bra mått på hur stor sannolikhet prognosen bedömer att det kommer bli en sjöbrisdag. Detta är användbart då sjöbrisens karaktärsdrag som alltid kommer gälla, men i olika storlek beroende

på signalstyrkan av sjöbrisen. Både vindhastigheten och vindstyrkan är direkt kopplade till hur seglare planerar sin strategi och taktik under dagen. Allt från hur man väljer att trimma båten för att maximera båtfarten till vilken sida av tävlingsområdet som kommer att kunna producera mest fördel med avsikt på vindriktning. Vindhastigheten spelar en extra stor roll i hur seglare lägger upp sin taktik, eftersom vinden beter sig annorlunda beroende på styrka när den interagerar med större fält av båtar på ett område.

Samtidigt finns det begränsningar i vilken grad av tillförlitlighet denna prognos kan användas och den bör därför inte ses som ett svar. Fortsatt bör prognosen jämföras med andra lokala prognoser för att utvärdera det meteorologiska händelseförloppet under tävlingsdagen. Osäkerheten verkar vara större när sjöbrissignalen är svagare än när sjöbrissignalen är starkare, vilket också bör tas i beaktande.

6 Slutsats

Syftet med arbetet var att bidra till ökad kunskap om sjöbrisens påverkan på vindförhållanden utanför San Pedros kust samt undersöka hur sjöbrisen kan modelleras för en operativ prognos. Målet var att utveckla ett maskinlärt prognosverktyg som kunde identifiera historiska sjöbrisfall, generera timvisa prognoser för vindhastighet och vindriktning samt uppskatta sjöbris sannolikheten.

En LSTM-modell utvecklades och jämfördes med vinddata från en NDBC-observationsstation. Resultatet visade att modellen kunde göra en bra approximation, speciellt i identifieringen av sjöbrisfall. Prognoserna gav något mindre fel och den bästa varianten var 1-timmesprognosen. Dock hade modellen svårare att uppskatta mer komplexa situationer och fall med svårare sjöbrissignaler.

Analysen visade att modellen kan komplettera GFS-prognoser och har därmed en operativ användbarhet för sjöbrisförhållanden i San Pedro. Däremot identifierades begränsningar i rådande geografisk indelning, träningsdata, komplext sjöbrisfall och klimatförändringar. För framtida utveckling skulle det kunna inkludera en finare upplösning på träningsdatan och fler observationspunkter i området.

Referenser

- [1] SMHI. *Sjöbris- en sval solig sommardag*. <https://www.smhi.se/kunskapsbanken/meteorologi/vind/en-sval-bris-soliga-sommardagar>. hämtad 6 februari 2026.
- [2] John E. Walsh. “Sea Breeze Theory and Applications”. I: *Journal of the Theory and Applications* 31 (1974).
- [3] S. T.K. Miller m.fl. “Sea breeze: Structure, forecasting, and impacts”. I: *Reviews of Geophysics* 41 (3 2003). ISSN: 87551209. DOI: 10.1029/2003RG000124.
- [4] Roland B. Stull. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 1988. DOI: 10.1007/978-94-009-3027-8.
- [5] Ytzhag Mahrer och Roger A. Pielke. “The Effects of Topography on Sea and Land Breezes in a Two-Dimensional Numerical Model”. I: *Monthly weather review* (mars 1977).
- [6] James R. Holton. *An Introduction to Dynamic Meteorology*. 4. utg. Vol. 88. International Geophysics Series. Elsevier Academic Press, 2004. ISBN: 0-12-354015-1.
- [7] Modern Physics. *Energy Cascade in Turbulence*. [Online]. Accessed: 16-Mar-2026. 2023. URL: <https://modern-physics.org/energy-cascade-in-turbulence/>.
- [8] A. B. C. Tijm, A. A. M. Holtslag och A. J. van Delden. “Observations and modeling of the sea breeze with the return current”. I: *Monthly Weather Review* 127 (maj 1999), s. 625–640. DOI: 10.1175/1520-0493(1999)127<0625:0AMOTS>2.0.CO;2.
- [9] Raymond W. Arritt. “Numerical modelling of the offshore extent of sea breezes”. I: *Quarterly Journal of the Royal Meteorological Society* 115 (april 1989), s. 547–570. DOI: <https://doi.org/10.1002/qj.49711548707>.
- [10] Dave Bergmann. *What is machine learning?* IBM Think. [Online]. Accessed: 16-Mar-2026. URL: <https://www.ibm.com/think/topics/machine-learning>.
- [11] IBM. *What is PyTorch?* IBM Think. [Online]. Accessed: 16-Mar-2026. URL: <https://www.ibm.com/think/topics/pytorch>.
- [12] PyTorch. *PyTorch Main Components*. PyTorch Documentation. [Online]. Accessed: 16-Mar-2026. URL: https://docs.pytorch.org/docs/stable/user_guide/pytorch_main_components.html.
- [13] Copernicus Climate Change Service (C3S). *ERA5 hourly data on single levels from 1940 to present*. [Online]. Accessed: 16-Mar-2026. 2018. DOI: 10.24381/cds.adbb2d47. URL: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels>.
- [14] National Centers for Environmental Information (NCEI). *Global Forecast System (GFS)*. [Online]. Accessed: 16-Mar-2026. 2026. URL: <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>.
- [15] National Centers for Environmental Prediction (NCEP). *NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive (Dataset d084001)*. UCAR Geoscience Data Exchange (GDEX). [Online]. Accessed: 16-Mar-2026. 2015. DOI: 10.5065/D65D8PWK. URL: <https://gdex.ucar.edu/datasets/d084001/>.
- [16] NCEP Environmental Modeling Center (EMC). *Global Forecast System (GFS) Documentation*. [Online]. Accessed: 16-Mar-2026. 2026. URL: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php.
- [17] Aki Lilja m.fl. *Survey on the Use of the Guide to Instruments and Methods of Observation (WMO-No. 8)*. Tekn. rapport Instruments and Observing Methods Report No. 135. World Meteorological Organization (WMO), 2021.